

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 18:49:18 ; Search time 1832.29 seconds  
(without alignments)  
- 11293.030 Million cell updates/sec

Title: US-09-105-117K-1\_COPY\_1016\_1726

Perfect score: 711

Sequence: 1 atggtgatcatggaatctt.....aactgatgttgatgggttag 711

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hlg.\*

3: gb\_in.\*

4: gb\_on.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	711	100.0	822	6	AX063771	AX063771 Sequence
2	711	100.0	822	6	AX244059	AX244059 Sequence
3	711	100.0	2374	1	CGLXSEG	X96471 C. glutamicu
4	711	100.0	2374	6	AP93933	AP93933 Sequence 2
5	711	100.0	333150	1	AP005277	AP005277 Coryneb
6	711	100.0	349980	6	AX127147	AX127147 Sequence
7	708	99.6	708	6	AX123539	AX123539 Sequence
8	317.6	44.7	1771	1	AB083133	AB083133 Coryneb
9	84.2	11.8	190050	1	AL646059	AL646059 Ralstonia
10	65.6	9.2	33285	1	SC5F8	AL357613 Streptomy
11	63.4	8.9	10542	1	AE004852	AE004852 Pseudomon
12	62.8	8.8	24336	1	AE008841	AE008841 Salmonell
13	62	8.7	10701	1	AE005522	AE005522 Escherich
14	62	8.7	286658	1	AP002563	AP002563 Escherich
15	61.6	8.7	4033	1	ASU65741	U65741 Aeromonas s
16	61.2	8.6	230050	1	AL627277	AL627277 Salmonell
17	60.4	8.5	636	6	AX030085	AX030085 Sequence
18	60.4	8.5	636	6	E49392	E49392 Process for
19	60.4	8.5	8029	1	ECFAPGK	AE000375 Escherich
20	60.4	8.5	10362	1	AE000375	AE000375 Escherich
21	60.4	8.5	141744	1	ECU28377	U28377 Escherichia
22	60	8.4	213050	1	AL646067	AL646067 Ralstonia
23	59.6	8.4	10747	1	AE012410	AE012410 Xanthomon
24	57	8.0	14844	1	MEU34849	U34849 Mycobacteri
25	57	8.0	14869	1	AE007056	AE007056 Mycobacte
26	57	8.0	15239	6	I86263	I86263 Sequence 17
27	57	8.0	38500	1	MTCY39	MTCY39 Mycobacteri
28	56.4	7.9	11139	1	AE009057	AE009057 Agrobacte
29	56.4	7.9	12305	1	AE008023	AE008023 Agrobacte
30	56.2	7.9	14029	1	AE006952	AE006952 Mycobacte
31	56.2	7.9	37218	1	MTCY2069	MTCY2069 Mycobacte
32	56	7.9	10333	1	AE013931	AE013931 Yersinia
33	56	7.9	208050	1	AJ414145	AJ414145 Yersinia
34	55.4	7.8	10393	1	AE011957	AE011957 Xanthomon
35	52.8	7.4	12070	1	AB011413	AB011413 Streptomy
36	52.8	7.4	14135	1	AE009566	AE009566 Brucella
37	50.2	7.1	10698	1	AE004134	AE004134 Vibrio ch
38	48.8	6.9	2878	1	AHAHYGENE	X89469 Aeromonas h
39	46	6.5	115666	2	AC105744	AC105744 Oryza sat
40	43.8	6.2	125020	9	AF429315	AF429315 Homo sapi
41	42.8	6.0	125020	9	AF429315	AF429315 Homo sapi
42	42.6	6.0	970	1	AB032557	AB032557 Streptomy
43	42.6	6.0	42210	1	SC1C2	AL031124 Streptomy
44	42.4	6.0	168398	2	AC097204	AC097204 Rattus no
45	42.2	5.9	149437	2	AP005610	AP005610 Oryza sat

#### ALIGNMENTS

RESULT 1  
AX063771

LOCUS

DEFINITION

AX063771

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AX063771 822 bp DNA linear PAT 24-JAN-2001  
Sequence 53 from Patent WO0100843.

AX063771 GI:12541483

Corynebacterium glutamicum.

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Corynebacteriaceae;

Corynebacterium.

1 (bases 1 to 822)

Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.

corynebacterium glutamicum genes encoding metabolic pathway

## proteins

JOURNAL Patent: WO 0100843-A 53 04-JAN-2001;  
BASF AKTIENGESSELLSCHAFT (DE)

## FEATURES

source  
1. .822  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"

CDS  
101..802  
/note="RXA01394"

/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC25127.1"

/db\_xref="GI:12541484"

/translation="MEIFITGLLLGASLLISGPQNVLVIKQIKREGLIALLVCLL

SDVFLFIAGTGLVDLLSNAPIVLDIMRWGGIAYLLFWAFAAKADMTNKVEAPQIIIE

ETEPVDDTPLGGSAVATDRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDA

FVFIQGAQYQDGTGRWIFAAAGFAASLIWFLPLVGFAGAAALSRPLSPKVRWVNVV

AVVMTALAIKMLMG"

BASE COUNT 167 a 192 c 246 g 217 t

ORIGIN

Query Match 100.0%; Score 711; DB 6; Length 822;

Best Local Similarity 100.0%; Pred. No. 2.5e-168;

Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGATCATCGAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60

DB 92 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 151

QY 61 ATCGACCGCAGAACTGCTGATTAACAAAGAAATTAAGCGCGAAGGACTCAATTGCG 120

DB 152 ATCGACCGCAGAACTGCTGATTAACAAAGAAATTAAGCGCGAAGGACTCAATTGCG 211

QY 121 GTTCTTCTCGTGTGTTAATTTCTACGCTCTTTTGTTCATCGCGGACCTTGGCGGTT 180

DB 212 GTTCTTCTCGTGTGTTAATTTCTACGCTCTTTTGTTCATCGCGGACCTTGGCGGTT 271

QY 181 GATCTTTTGTCCAAATCGCGCGGATCTGCTCGATATTATGCGTGGGGTGCATCGCT 240

DB 272 GATCTTTTGTCCAAATCGCGCGGATCTGCTCGATATTATGCGTGGGGTGCATCGCT 331

QY 241 TACCTGTTATGTTTGGCGTCTATGCGAGAAAGCGCATGACAAACAAAGGTGGAAGCG 300

DB 332 TACCTGTTATGTTTGGCGTCTATGCGAGAAAGCGCATGACAAACAAAGGTGGAAGCG 391

QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCCCTTTGGCGGTTG 360

DB 392 CCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCCCTTTGGCGGTTG 451

QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 420

DB 452 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 511

QY 421 GTTTCGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGTTGAAACCCGAATCGGTAT 480

DB 512 GTTTCGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGTTGAAACCCGAATCGGTAT 571

QY 481 TTGGACGCGTTTGTGTTATCGCGCGCGTCCGCGCGCAATACGGCGAACACCGGACGGTGG 540

DB 572 TTGGACGCGTTTGTGTTATCGCGCGCGTCCGCGCGCAATACGGCGAACACCGGACGGTGG 631

QY 541 ATTTTCGCGCGTGGCGGTTCCGCGCAACCGTGTGTTGTTCCCGCTGTTGGGTTTCGGC 600

DB 632 ATTTTCGCGCGTGGCGGTTCCGCGCAACCGTGTGTTGTTCCCGCTGTTGGGTTTCGGC 691

QY 601 GCAGCAGCATTTCTACGCGCGCTGTCCAGCCCAAGGTGTGGCGTGTGATCAACGTCGTC 660

DB 692 GCAGCAGCATTTCTACGCGCGCTGTCCAGCCCAAGGTGTGGCGTGTGATCAACGTCGTC 751

QY 661 GTGGCAGTTGTGATACCGCGATTTGGCCATCAAACTGATGTTGATGGGTTAG 711

DB 752 GTGGCAGTTGTGATACCGCGATTTGGCCATCAAACTGATGTTGATGGGTTAG 802

## RESULT 2

AX244059

LOCUS

DEFINITION

AX244059

ACCESSION

AX244059

VERSION

AX244059.1

KEYWORDS

ORGANISM

Corynebacterium glutamicum.

Corynebacterium glutamicum

Bacteria; Firmicutes; Actinobacteri

Actinomycetales; Corynebacteri

Corynebacterium.

1 (bases 1 to 822)

Pompejus, M., Kroege, B., Schroeder, H., Zelder, O., Haberhauer, G.,

Kim, J.W., Lee, H.S. and Hwang, B.J.

Corynebacterium glutamicum genes encoding metabolic pathway

proteins

Patent: WO 0166573-A 51 13-SEP-2001;

BASF AKTIENGESSELLSCHAFT (DE)

FEATURES

source

1. .822

/organism="Corynebacterium glutamicum"

/db\_xref="taxon:1718"

101..802

/note="RXA01394"

/codon\_start=1

/transl\_table=11

/protein\_id="CAC88508.1"

/db\_xref="GI:15859124"

/translation="MEIFITGLLLGASLLISGPQNVLVIKQIKREGLIALLVCLL

SDVFLFIAGTGLVDLLSNAPIVLDIMRWGGIAYLLFWAFAAKADMTNKVEAPQIIIE

ETEPVDDTPLGGSAVATDRNRVRVESVDKQVWVKPMLMAIVLTWLNPNAYLDA

FVFIQGAQYQDGTGRWIFAAAGFAASLIWFLPLVGFAGAAALSRPLSPKVRWVNVV

AVVMTALAIKMLMG"

BASE COUNT 167 a 192 c 246 g 217 t

ORIGIN

Query Match 100.0%; Score 711; DB 6; Length 822;

Best Local Similarity 100.0%; Pred. No. 2.5e-168;

Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGATCATCGAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60

DB 92 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 151

QY 61 ATCGACCGCAGAACTGCTGATTAACAAAGAAATTAAGCGCGAAGGACTCAATTGCG 120

DB 152 ATCGACCGCAGAACTGCTGATTAACAAAGAAATTAAGCGCGAAGGACTCAATTGCG 211

QY 121 GTTCTTCTCGTGTGTTAATTTCTACGCTCTTTTGTTCATCGCGGACCTTGGCGGTT 180

DB 212 GTTCTTCTCGTGTGTTAATTTCTACGCTCTTTTGTTCATCGCGGACCTTGGCGGTT 271

QY 181 GATCTTTTGTCCAAATCGCGCGGATCTGCTCGATATTATGCGTGGGGTGCATCGCT 240

DB 272 GATCTTTTGTCCAAATCGCGCGGATCTGCTCGATATTATGCGTGGGGTGCATCGCT 331

QY 241 TACCTGTTATGTTTGGCGTCTATGCGAGAAAGCGCATGACAAACAAAGGTGGAAGCG 300

DB 332 TACCTGTTATGTTTGGCGTCTATGCGAGAAAGCGCATGACAAACAAAGGTGGAAGCG 391

QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCCCTTTGGCGGTTG 360

DB 392 CCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCCCTTTGGCGGTTG 451

QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 420

DB 452 GCGGTGGCCACTGACACGCGCAACCGGTGCGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 511

QY 421 GTTTCGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGTTGAAACCCGAATCGGTAT 480

DB 512 GTTTCGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGTTGAAACCCGAATCGGTAT 571

QY 481 TTGGACGCGTTTGTGTTATCGCGCGCGTCCGCGCGCAATACGGCGAACACCGGACGGTGG 540

DB 572 TTGGACGCGTTTGTGTTATCGCGCGCGTCCGCGCGCAATACGGCGAACACCGGACGGTGG 631

QY 541 ATTTTCGCGCGTGGCGGTTCCGCGCAACCGTGTGTTGTTCCCGCTGTTGGGTTTCGGC 600

DB 632 ATTTTCGCGCGTGGCGGTTCCGCGCAACCGTGTGTTGTTCCCGCTGTTGGGTTTCGGC 691

QY 601 GCAGCAGCATTTCTACGCGCGCTGTCCAGCCCAAGGTGTGGCGTGTGATCAACGTCGTC 660

DB 692 GCAGCAGCATTTCTACGCGCGCTGTCCAGCCCAAGGTGTGGCGTGTGATCAACGTCGTC 751

QY 661 GTGGCAGTTGTGATACCGCGATTTGGCCATCAAACTGATGTTGATGGGTTAG 711

DB 752 GTGGCAGTTGTGATACCGCGATTTGGCCATCAAACTGATGTTGATGGGTTAG 802

```

QY 481 TTGACGCGTTTGTGTTATCGCGCGTTCGGCGCAATACGGCGACACCGGACGGTGG 540
|
|
|
Db 572 TTGACGCGTTTGTGTTATCGCGCGTTCGGCGCAATACGGCGACACCGGACGGTGG 631
|
|
|
QY 541 ATTTTCGCGCTCGCGGTTTCGGCGAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 600
|
|
|
Db 632 ATTTTCGCGCTCGCGGTTTCGGCGAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 691
|
|
|
QY 601 GCAGCAGCATTGTCACGCGCGTTCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 660
|
|
|
Db 692 GCAGCAGCATTGTCACGCGCGTTCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 751
|
|
|
QY 661 GTGGCAGTTGTGATGACCGCATTGGCCATCAACTGATGTTGATGGGTTAG 711
|
|
|
Db 752 GTGGCAGTTGTGATGACCGCATTGGCCATCAACTGATGTTGATGGGTTAG 802
|
|
|

RESULT 3
CGLYSEG
LOCUS
DEFINITION
ACCESSION X96471
VERSION X96471.1 GI:1729753
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine exporter protein; Lysine governor.
SOURCE Corynebacterium glutamicum.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE 1 (bases 1 to 2374)
AUTHORS Vrljic.M., Sahm.H. and Eggeling,L.
TITLE A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum
JOURNAL Mol. Microbiol. 22 (5), 815-826 (1996)
MEDLINE 97126810
PUBMED 8971704
REFERENCE 2 (bases 1 to 2374)
AUTHORS Vrljic.M.M.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie 1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG
FEATURES
source
/organism="Corynebacterium glutamicum"
/strain="R127"
/db_xref="taxon:1718"
complement (82..954)
/gene="lysG"
complement (82..954)
/gene="lysG"
/note="lysine governor"
/codon_start=1
/transl_table=11
/product="lysine export regulator protein"
/protein_id="CAA65323.1"
/db_xref="GI:1729754"
/db_xref="SWISS-PROT:P94632"
/translation="MNPQLDLSIDEGSFEGASIALSISPSAVSORVKALBHHVG RVLSRTPAKATAGEVLVQARMKVLLQATKQLSGLAEIPLTIAINADSLSTW PFPVNEVASGGATLRLLEDLISLLRQAVTRAEANPVAGCEVLELGTWR HLATATSLRAYVVDGKLDWAAMPVLRFGPKVDLDRDLGRVDGPGVRRRVISPS AEGFGEAIRRGILGWLIPETQAPMLKAGEVILLDEIPIDTPMTWQRWRESRLARL TDVVDAAIEGLRP"
1025..1726
/gene="lysE"
1025..1726
/gene="lysE"
/codon_start=1
/transl_table=11
/product="Lysine exporter protein"
/protein_id="CAA65324.2"

```

```

/db_xref="GI:13397387"
/db_xref="SWISS-PROT:P94633"
/translation="MEIFITGLLGLSLLSISGPQNVLVLKQIKREGLIALLVCLL SDVFLPFIAGTLGVDLLSNAAPVLDIMRWGGIATLLFWFAAKADNTNKVEAPQILE ETEPVPDDTPLGGSVATDTRNRVRVEVDKORVWVKPMLMAIIVLTWLNPNAYLDA FVTIGVGQAQYGDTRWIFPAAGAFASLIWFLPLVGFGAALSRPLSPKVRWIVVV AVVMTALAIKMLMG"
complement (1723..2352)
/gene="orf3"
complement (1723..2352)
/gene="orf3"
/codon_start=1
/transl_table=11
/protein_id="CAA65325.1"
/db_xref="GI:1729756"
/db_xref="SPTREMBL:P94634"
/translation="MYALRDIVASGKALYVIGSYGPETLAEAAEFMAEBCPLLIHQ PSYIINRWVEEFDGDNLLQSAANNGLVIAFSPLAQLLTKYLDGIPGSRASQ GKSLSEGMLNVNNDIMVRKLDIAQRGQSLAQWALAWLRQEGYGAADVTSALIGA SSVEQLDNLSDLSLNLEFSDAELEAIDEISHDAGINIWAKATDSKTREN"
BASE COUNT 526 a 640 c 648 g 560 t
ORIGIN
Query Match 100.0%; Score 711; DB 1; Length 2374;
Best Local Similarity 100.0%; Pred. No. 2.8e-168;
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGTGATCATGGAAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
|
|
|
Db 1016 ATGGTGATCATGGAAATCTTCATTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 1075
|
|
|
QY 61 ATCGACCGCAGATGACTCGTGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 120
|
|
|
Db 1076 ATCGACCGCAGATGACTCGTGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 1135
|
|
|
QY 121 GTTCTTCTCGTGTGTTAAATTTCTGACGTCTTTTGTTCATCGCGGCACCTTCGGGGTT 180
|
|
|
Db 1136 GTTCTTCTCGTGTGTTAAATTTCTGACGTCTTTTGTTCATCGCGGCACCTTCGGGGTT 1195
|
|
|
QY 181 GATCTTTTGCCAAATGCGGCGCATCGTGTGATTAATATGCTGGGGTGGATCGCT 240
|
|
|
Db 1196 GATCTTTTGCCAAATGCGGCGCATCGTGTGATTAATATGCTGGGGTGGATCGCT 1255
|
|
|
QY 241 TACCTGTTATGTTTGGCTTCATGCGCAGCAAGAACCCCATGACAAACAGGTGGAAGCG 300
|
|
|
Db 1256 TACCTGTTATGTTTGGCTTCATGCGCAGCAAGAACCCCATGACAAACAGGTGGAAGCG 1315
|
|
|
QY 301 CCACAGATCATTTGAAGAAACAGAACCCGTCGCCGATGACACGCTTTTGGGCGGTTGCG 360
|
|
|
Db 1316 CCACAGATCATTTGAAGAAACAGAACCCGTCGCCGATGACACGCTTTTGGGCGGTTGCG 1375
|
|
|
QY 361 CGGTTGGCACTGACACGCGCAACCGGTTGCGGTTGAGGTGAGCTGATGATGACGCGG 420
|
|
|
Db 1376 CGGTTGGCACTGACACGCGCAACCGGTTGCGGTTGAGGTGAGCTGATGATGACGCGG 1435
|
|
|
QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGACCCGAAATGCGTAT 480
|
|
|
Db 1436 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAAACCCGAAATGCGTAT 1495
|
|
|
QY 481 TTGACGCGTTTGTGTTTATCGCGCGCTCGCGCGCAATACCGCGCAACCGGACGGTGG 540
|
|
|
Db 1496 TTGACGCGTTTGTGTTTATCGCGCGCTCGCGCGCAATACCGCGCAACCGGACGGTGG 1555
|
|
|
QY 541 ATTTTCGCGCTCGCGGTTTCGGCGAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 600
|
|
|
Db 1556 ATTTTCGCGCTCGCGGTTTCGGCGAAGCTGATCTGGTTCCCGCTGGTGGTTTCGGC 1615
|
|
|
QY 601 GCAGCAGCATTGTCACGCGCGTTCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 660
|
|
|
Db 1616 GCAGCAGCATTGTCACGCGCGTTCAGCCCAAGTGTGGCGTGGATCAACGTCGTC 1675
|
|
|
QY 661 GTGGCAGTTGTGATGACCGCATTGGCCATCAACTGATGTTGATGGGTTAG 711
|
|
|

```

Db 1676 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGTTAG 1726

RESULT 4  
LOCUS A93933 2374 bp DNA linear PAT 22-JAN-2000  
DEFINITION Sequence 2 from Patent WO9723597.  
ACCESSION A93933  
VERSION A93933.1 GI:6742037  
KEYWORDS  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
1 (bases 1 to 2374)  
REFERENCE Vrljic.M. and Eggeing.L.  
AUTHORS PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED  
TITLE ACTIVITY OF EXPORT CARRIERS  
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;  
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLJIC MARINA (DE)  
FEATURES  
source Location/Qualifiers  
1 .2374  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
BASE COUNT 526 a 640 c 648 g 560 t  
ORIGIN  
Query Match 100.0%; Score 711; DB 6; Length 2374;  
Best Local Similarity 100.0%; Pred. No. 2.8e-168;  
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGTGATCATGGAATCTTCAATACAGTCTGCTTTGGGGCCAGCTTTTACTGTCC 60  
Db 1016 ATGGTGATCATGGAATCTTCAATACAGTCTGCTTTGGGGCCAGCTTTTACTGTCC 1075  
QY 61 ATCGGACCCAGAACTGCTGCTGATTAACAAGGAATTAACGCGGAAGGACTCAATCGG 120  
Db 1076 ATCGGACCCAGAACTGCTGCTGATTAACAAGGAATTAACGCGGAAGGACTCAATCGG 1135  
QY 121 GTTCTTCTGCTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGCGGTT 180  
Db 1136 GTTCTTCTGCTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGCGGTT 1195  
QY 181 GATCTTTTGTCAATGCGCGCGATCTGCTCGATATATCGCTGCGGTGGATCGCT 240  
Db 1196 GATCTTTTGTCAATGCGCGCGATCTGCTCGATATATCGCTGCGGTGGATCGCT 1255  
QY 241 TACCTGTTATGTTTCCGCTCATGCGAGGAAGACGCCATGACAAACAAGTGAAGCG 300  
Db 1256 TACCTGTTATGTTTCCGCTCATGCGAGGAAGACGCCATGACAAACAAGTGAAGCG 1315  
QY 301 CCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTTCG 360  
Db 1316 CCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTTCG 1375  
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTCGCGGTGGAGGTGAGCGTGAAGCAGCGG 420  
Db 1376 GCGGTGGCCACTGACACGCGCAACCGGTCGCGGTGGAGGTGAGCGTGAAGCAGCGG 1435  
QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAAACCCGAATCGGTAT 480  
Db 1436 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAAACCCGAATCGGTAT 1495  
QY 481 TTGAGCGGTTTGTGTTTATCGCGCGCGTTCGCGCGCAATACCGCGACACCGGCGGTGG 540  
Db 1496 TTGAGCGGTTTGTGTTTATCGCGCGCGTTCGCGCGCAATACCGCGACACCGGCGGTGG 1555  
QY 541 ATTTTTCGCGCTGCGCGCTTCGCGCAAGCTGATCTGTTCCCGTGGTGGGTTTCGCG 600  
Db 1556 ATTTTTCGCGCTGCGCGCTTCGCGCAAGCTGATCTGTTCCCGTGGTGGGTTTCGCG 1615  
QY 601 GCAGCAGCATTTGTACGCGCCCGCTGTCCAGCCCAAGGTGTGGCGTGGATCAACGTCGTC 660

Db 1616 GCACGACATTTGACCCCGCTGTCCAGCCCAAGGTGTGGCGTGCATCAACGTCGTC 1675

QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGTTAG 711

Db 1676 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGTTAG 1726

RESULT 5  
LOCUS AP005277/c 333150 bp DNA linear BCT 08-AUG-2002  
DEFINITION Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 4/10  
ACCESSION AP005277 BA000036  
VERSION AP005277.1 GI:21323710  
KEYWORDS  
SOURCE Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.  
ORGANISM Corynebacterium glutamicum ATCC 13032  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
1  
REFERENCE Nakagawa,S.  
AUTHORS Complete genomic sequence of Corynebacterium glutamicum ATCC 13032  
TITLE Unpublished  
JOURNAL  
REFERENCE 2 (bases 1 to 333150)  
AUTHORS Nakagawa,S.  
TITLE Direct Submission  
JOURNAL  
COMMENT This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.  
FEATURES  
source Location/Qualifiers  
1 .333150  
/organism="Corynebacterium glutamicum ATCC 13032"  
/strain="ATCC 13032"  
/db\_xref="taxon:196627"  
/notes="ATCC 13032"  
complement (63..1292)  
/gene="Cgl10944"  
complement (63..1292)  
/gene="Cgl10944"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical membrane protein"  
/protein\_id="BAB98337.1"  
/db\_xref="GI:21323711"  
/translation="MTNPTENRRLIHWGLQNTQDQIVAAKTIVLPWLLQAAGP  
FLALVPIREAGMLPQRAITGWLRSKRVVIGNSQFVSAIGVAAALFLRG  
WALGITVI VLLAALSIFRSMCSIASKDVQKRVISKRGVLTGRATVIGVGMVLVAGL  
AIAFLGSHSPTRVLAAVTASFWLFAIVFARIEYAKPATPKNAPSANPWRRCI  
AALDDKAFRRFVLVRSMLVTALSTAFI VALAASGNSIDSLGFLFIAGSLAVMGV  
RISGIWSHDSKNVMAGGALFGSIVILVLSAFAPAQINTLVFLPSFLITLAHTA  
IRVARTVYMDMAEGDQRTFYVADANTLMGVLLIIVGALSGLFIAPNEAALLFLAI  
GLGITSARGLKEVSAG"  
1382..2917  
/gene="Cgl10945"  
1382..2917  
/gene="Cgl10945"  
/codon\_start=1  
/transl\_table=11  
/product="Putative multicopper oxidases"  
/protein\_id="BAB98338.1"  
/db\_xref="GI:21323712"  
/translation="MPENLRRTFFKAGVLAATVGAQVLVACSSDDVRVYGGEPRTL  
PIPADLGTREGSSVHFALFAEQESQILPDVTTKTGNGTHLPTLVVKGDDVHV  
DVINLDEMTHVHGMKLPALADGPHSPFGQTSPTVTVANDAAATLWYRPHTHG  
LTGLHAYRGLAGMLIIVEDEATDKLDLPREYGVDDI DPLVLMDRHFLPDGSLDEBDLP  
GLGDTTANGITNAHFATRRVRFLVNGSNRMFLNLA FSDTRTFQVDFSGSLD  
EPQDRTTLAIPGERWEI VVELEPDEDTLVESVGFEDYGVDPDEFVDFPGMSDSFOL  
LTTGPSDDAAQAAPALPGVLVKSTEPDVIDATERTIMNTFSINDLQMDMQRVDDVID  
HDQPEWIVTNDNSDPHNFVHDAREKVLKFESTDVELFENDGMDKTGVLPFGATATL





QY 61 ATCGACCGCAGAAATGTAAGTATTAACCAAGAAATTAAGCGAAGGACTCATTGCG 120  
 |||||  
 Db 328943 ATCGACCGCAGAAATGTAAGTATTAACCAAGAAATTAAGCGAAGGACTCATTGCG 328884  
 QY 121 GTTCTTCTGCTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGCGACCTTTGGGCGTT 180  
 |||||  
 Db 328883 GTTCTTCTGCTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGCGACCTTTGGGCGTT 328824  
 QY 181 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATTAATCGCTGGGGTGGCATCGCT 240  
 |||||  
 Db 328823 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATTAATCGCTGGGGTGGCATCGCT 328764  
 QY 241 TACCTGTTATGTTTGGCTGATGCGAAGAGCGCATGACAAACAAGTGAAGCG 300  
 |||||  
 Db 328763 TACCTGTTATGTTTGGCTGATGCGAAGAGAGCGCATGACAAACAAGTGAAGCG 328704  
 QY 301 CCACAGATCATTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTGCG 360  
 |||||  
 Db 328703 CCACAGATCATTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTGCG 328644  
 QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGGTGAGCTGATGAAGCAGCGG 420  
 |||||  
 Db 328643 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGGTGAGCTGATGAAGCAGCGG 328584  
 QY 421 GTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGGAATCGGTAT 480  
 |||||  
 Db 328583 GTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGGAATCGGTAT 328524  
 QY 481 TTGGACGCGTGTGTTTATCGCGCGCGTCCGCGGCGCAATACCGCGACACCGGACGTTGG 540  
 |||||  
 Db 328523 TTGGACGCGTGTGTTTATCGCGCGCGTCCGCGGCGCAATACCGCGACACCGGACGTTGG 328464  
 QY 541 ATTTTCGCGTGGCGGCTTCCGCGCAAGCTGATCTGTTCCCTCGTGGTGGTTTCGCG 600  
 |||||  
 Db 328463 ATTTTCGCGTGGCGGCTTCCGCGCAAGCTGATCTGTTCCCTCGTGGTGGTTTCGCG 328404  
 QY 601 GCAGCAGCATTTGACGCGCGCTGCCAGCCCAAGGTGTCGCTGATCAACGTCGTC 660  
 |||||  
 Db 328403 GCAGCAGCATTTGACGCGCGCTGCCAGCCCAAGGTGTCGCTGATCAACGTCGTC 328344  
 QY 661 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 711  
 |||||  
 Db 328343 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 328293

RESULT 6  
 AX127147/c AX127147 34980 bp DNA linear PAT 11-MAY-2001  
 LOCUS  
 DEFINITION Sequence 7063 from Patent EP1108790.  
 ACCESSION AX127147 AX114121  
 VERSION AX127147.1 GI:14041135  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Corynebacterium glutamicum.  
 Corynebacterium glutamicum  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 Corynebacterium.

REFERENCE 1 (bases 1 to 349980)  
 AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
 Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
 TITLE Novel polynucleotides  
 JOURNAL Patent: EP 1108790-A 7063 20-JUN-2001;  
 KYOWA HAKKO KOGYO CO., LTD. (JP)  
 FEATURES  
 Location/Qualifiers  
 source  
 1. 349980  
 /organism="Corynebacterium glutamicum"  
 /db\_xref="taxon:1718"  
 /note="Seq 1 to long (3.309.400) split in 11, seq 7063  
 1.200.001 1.549.980"  
 80727 a 91049 c 97618 g 80586 t

BASE COUNT 80727 a 91049 c 97618 g 80586 t  
 ORIGIN

Query Match 100.0%; Score 711; DB 6; Length 349980;

Best Local Similarity 100.0%; Pred. No. 5.5e-168;  
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGTGTATCATGGAATCTTCTTACATACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCC 60  
 |||||  
 Db 128953 ATGTGTATCATGGAATCTTCTTACATACAGGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCC 128894  
 QY 61 ATCGGACCGCAGAAATGTAAGTATTAACCAAGAAATTAAGCGCGAAGGACTCATTGCG 120  
 |||||  
 Db 128893 ATCGGACCGCAGAAATGTAAGTATTAACCAAGAAATTAAGCGCGAAGGACTCATTGCG 128834  
 QY 121 GTTCTTCTGCTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGCGACCTTTGGGCGTT 180  
 |||||  
 Db 128833 GTTCTTCTGCTGTTAAATTTCTGACGCTCTTTTGTTCATCGCCGCGACCTTTGGGCGTT 128774  
 QY 181 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATTAATCGCTGGGGTGGCATCGCT 240  
 |||||  
 Db 128773 GATCTTTTGTTCATGCGCGCGGATCGTGTGCTGATATTAATCGCTGGGGTGGCATCGCT 128714  
 QY 241 TACCTGTTATGTTTGGCTGATGCGAAGAGCGCATGACAAACAAGTGAAGCG 300  
 |||||  
 Db 128713 TACCTGTTATGTTTGGCTGATGCGAAGAGCGCATGACAAACAAGTGAAGCG 128654  
 QY 301 CCACAGATCATTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTGCG 360  
 |||||  
 Db 128653 CCACAGATCATTGAAGAAACAGAACCAACCGTGCCCGATGACACGCCCTTTGGGCGGTTGCG 128594  
 QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGGTGAGCTGATGAAGCAGCGG 420  
 |||||  
 Db 128593 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGGTGAGCTGATGAAGCAGCGG 128534  
 QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGGAATCGGTAT 480  
 |||||  
 Db 128533 GTTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGGAATCGGTAT 128474  
 QY 481 TTGGACGCGTGTGTTTATCGCGCGCGTCCGCGGCGCAATACCGCGACACCGGACGTTGG 540  
 |||||  
 Db 128473 TTGGACGCGTGTGTTTATCGCGCGCGTCCGCGGCGCAATACCGCGACACCGGACGTTGG 128414  
 QY 541 ATTTTCGCGTGGCGGCTTCCGCGCAAGCTGATCTGTTTCCCGTGGTGGGTTTCGCG 600  
 |||||  
 Db 128413 ATTTTCGCGTGGCGGCTTCCGCGCAAGCTGATCTGTTTCCCGTGGTGGGTTTCGCG 128354  
 QY 601 GCAGCAGCATTTGACGCGCGCTGCCAGCCCAAGGTGTCGCTGATCAACGTCGTC 660  
 |||||  
 Db 128353 GCAGCAGCATTTGACGCGCGCTGCCAGCCCAAGGTGTCGCTGATCAACGTCGTC 128294  
 QY 661 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 711  
 |||||  
 Db 128293 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 128243

RESULT 7  
 AX123539 AX123539 708 bp DNA linear PAT 11-MAY-2001  
 LOCUS  
 DEFINITION Sequence 3455 from Patent EP1108790.  
 ACCESSION AX123539  
 VERSION AX123539.1 GI:14041027  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Corynebacterium glutamicum.  
 Corynebacterium glutamicum  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 Corynebacterium.

REFERENCE 1 (bases 1 to 708)  
 AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
 Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
 TITLE Novel polynucleotides  
 JOURNAL Patent: EP 1108790-A 3455 20-JUN-2001;  
 KYOWA HAKKO KOGYO CO., LTD. (JP)  
 FEATURES  
 Location/Qualifiers  
 source  
 1. 708  
 /organism="Corynebacterium glutamicum"

```
BASE COUNT      134 a      173 c      221 g      180 t
ORIGIN
Query Match      99.6%; Score 708; DB 6; Length 708;
Best Local Similarity 100.0%; Pred. No. 1.4e-167;
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
DB 1 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60
QY 61 ATCGGACCGAGAAATGCTAGTGTGATTAACAAGGAATTAAGCGGAGAGTCAATTGCG 120
DB 61 ATCGGACCGAGAAATGCTAGTGTGATTAACAAGGAATTAAGCGGAGAGTCAATTGCG 120
QY 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTGGCGTT 180
DB 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTGGCGTT 180
QY 181 GATCTTTTGTCCAAATCGCGCGGATCGTGTCTGATTAATATGCTGGGTGGCATCGCT 240
DB 181 GATCTTTTGTCCAAATCGCGCGGATCGTGTCTGATTAATATGCTGGGTGGCATCGCT 240
QY 241 TACCTGTTATGTTTCCGCTCATGGCAGGAAAGCGCATGACAAACAAGTGAAGCG 300
DB 241 TACCTGTTATGTTTCCGCTCATGGCAGGAAAGCGCATGACAAACAAGTGAAGCG 300
QY 301 CCACAGATCATTCAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTGG 360
DB 301 CCACAGATCATTCAGAAACAGAACCAACCGTCCCGATGACACGCTTTGGCGGTTGG 360
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGTGGAGTGCATTAAGCAGCG 420
DB 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGGAGTGGAGTGCATTAAGCAGCG 420
QY 421 GTTGGGTTAAAGCCCATGTTGATGCAATCGTGTGCTGACCTGTTGAACCCGAATCGTAT 480
DB 421 GTTGGGTTAAAGCCCATGTTGATGCAATCGTGTGCTGACCTGTTGAACCCGAATCGTAT 480
QY 481 TTGGACGCTTTGTTTATCGCGCGCTGCGGCGCAATACGGCGACACCGGAGTGG 540
DB 481 TTGGACGCTTTGTTTATCGCGCGCTGCGGCGCAATACGGCGACACCGGAGTGG 540
QY 541 ATTTTCGCGCTGGCGGCTTCGCGCAAGCTGATCTGTTCCCGTGGTGGTTTCGCG 600
DB 541 ATTTTCGCGCTGGCGGCTTCGCGCAAGCTGATCTGTTCCCGTGGTGGTTTCGCG 600
QY 601 GCAGCAGATGTCACGCGCGCTGTCACGCCCAAGGTGCGGTGGATCAACGTCGTC 660
DB 601 GCAGCAGATGTCACGCGCGCTGTCACGCCCAAGGTGCGGTGGATCAACGTCGTC 660
QY 661 GTGGCAGTTGTCATGACCGCATGGCCATCAAACTGATGTTGATGGT 708
DB 661 GTGGCAGTTGTCATGACCGCATGGCCATCAAACTGATGTTGATGGT 708

RESULT 8
AB083133      1771 bp      DNA      linear      BCT 06-APR-2002
LOCUS      Corynebacterium efficiens lysG, lysE genes for Lysine export
DEFINITION      transcriptional regulatory protein, Lysine exporter protein,
complete cds.
ACCESSION      AB083133
VERSION      AB083133.1 GI:20065738
KEYWORDS
SOURCE      Corynebacterium efficiens DNA.
ORGANISM      Corynebacterium efficiens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE      1 Itaya,H., Kimura,E., Kawahara,Y. and Sugimoto,S.
AUTHORS      lysG, lysE of Corynebacterium efficiens
TITLE
```

## JOURNAL

Published Only in Database (2002)

## REFERENCE

2 (bases 1 to 1771)

## AUTHORS

Itaya,H., Kimura,E., Kawahara,Y. and Sugimoto,S.

## TITLE

Direct Submission

## JOURNAL

Submitted (04-APR-2002) Hiroshi Itaya, AJINOMOTO CO., INC,  
Fermentation & Biotechnology Laboratories; 1-1, Suzuki-Cho,  
Kawasaki-Ku, Kawasaki, Kanagawa 210-8681, Japan  
(E-mail:hiroshi.itaya@ajinomoto.com, Tel:81-44-244-7123 (ex.4146),  
Fax:81-44-222-0129)

## FEATURES

Location/Qualifiers

1. 1771

/organism="Corynebacterium efficiens"

/db\_xref="taxon:152794"

complement(47..919)

/genes="lysG"

complement(47..919)

/genes="lysG"

/codon\_start=1

/transl\_table=11

/product="Lysine export transcriptional regulatory

protein"

/protein\_id="BAB88826.1"

/db\_xref="GI:20065739"

/translation="MNPILHLDLTLTIIDEGSFENASLALSISPSAVSQRIKALEKSVG  
RVLSRTQPAVATEAGEVLEQARQKMLAQETREQLABRLDEIPTVAINADSLSTW  
PPVFAEVAHMGAVTLTLRVEDRAHTLSLLRRGSVLGAVTREADPVAGCEVRLGVNR  
HLPVATPELRAVTVDPQDPVEMPVLRSPNDVLODRLEGRVDCGAVARRVSVVPS  
AEGFGAERVLGLGWLIPPEAQAPMLAAGDVQLDEKVVDTPLYWQRWRLSRLARL  
TDAVDAARAGLRT"

981..1676

/genes="lysE"

981..1676

/genes="lysE"

/codon\_start=1

/transl\_table=11

/product="Lysine exporter protein"

/protein\_id="BAB88827.1"

/db\_xref="GI:20065740"

/translation="MRDMELFVTLGLLGSALLAIGPQNVLVIKQIKREGITAVLIV  
CLSDVVLFTGLTGLVGLSDTAPIILDIRMCGIAYLLWFVFAVMAARDALARTVTF  
VHSEPVAAASGGVTTKQRPRLRITSGTQVWVPMMLAIVLTWLPNAYLDAFV  
FIGVGQAQYGETGRWTFFAAGAFASLVWFLVGYGAALSRPLSSPRVWRWINIGAV  
VLTGLAVKILMG"

BASE COUNT

294 a 603 c 545 g 329 t

## ORIGIN

Query Match 44.7%; Score 317.6; DB 1; Length 1771;

Best Local Similarity 67.7%; Pred. No. 3.1e-69;

Matches 481; Conservative 0; Mismatches 214; Indels 15; Gaps 2;

QY 1 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60

DB 981 ATCGGACCATGGAATCTTTGTACCGGTTGTTGTTGGAGCCAGTCTGCTGTGTGGCC 1040

QY 61 ATCGGACCGAGAAATGCTAGTGTGATTAACAAGGAATTAAGCGGAGAGTCAATTGCG 120

DB 1041 ATCGGACCATGGAATGCTGCTGATCAACAGGCGATCAACGCGGAGTCAACGCGCC 1100

QY 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTGGCGTT 180

DB 1101 GTCATCATGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1160

QY 181 GATCTTTTGTCCAAATCGCGCGGATCGTGTCTGATTAATATGCTGGGTGGCATCGCT 240

DB 1161 GGCCTGATCTCCGACACCGCGCGGATCATTTCTGACATCTGCTGCTGCTGCTGCTGCTG 1220

QY 241 TACCTGTTATGTTTGGCTGTCATGCGGAGCGAAGAGCCCATGACAAAGGTGGAAGCG 300

DB 1221 TACCTGCTGTTGTTGCGGTGATGCGGCGCGCGGAGACGCCCTCGCGGCCCGCAGAGGTA 1280

QY 301 CCACAGATCATTCAGAAACAGAACCAACCGTCCCGCATGACACGCTTTGGCGGTTGG 360

DB 1281 ACCTTTGTC-----GAGCATTCGGAACCCGTTG---CGGACGCGTCCGCTCC 1325



```

/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE PROTEIN"
/protein_id="CAD13913.1"
/db_xref="GI:17427395"
/translation="MTGPGSPAPLIGALSFGFLGASLIIVAIGQNAAYLVRLQGLRRE
YLVGVLLCALCDMALIALGVAGMTGLISAPPAWLTAVRWAGAFLILAYGARAFRAAW
RGAERLQARNEDKASHAQVLAASALALSILNPHVYLDITVLLCAIGGRVAMPANVAFAG
GAMCASILWFLSLGFGARLLRPVPARPVAVNRVLDALIGAVMVAIALTILMG"
complement(2924..3910)
/gene="ftsY"
/ncbi="RSC0386; RS03354"
complement(2924..3910)
/gene="ftsY"
/function="cell processes; cell division"
/note="Product confidence : putative
Gene name confidence : putative
Predicted by Codon usage
predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE CELL DIVISION PROTEIN"
/protein_id="CAD13914.1"
/db_xref="GI:17427396"
/translation="WPDTPAAPALDQADDIETVTPPVVEQARKWMSRLRSLGSK
TSKSLTILFVGKVDLEALFELEETALLMADAGVDATEYLLDELRRVKAQRIETAEVG
KIALRDLLEHLEFLEKTVIGRDOPTVIMTAGVNGAGKTSISGLCKHFQYTGQSVL
LAAGDTFAARAEQVLVINGRNVTVAQESGDPAAVIFDVAARARGIDIVNADTA
GRILPTQLHMEELKVRVIGKAMATAPHETLLVIDANTGONALAQVAFDALLGLTG
LIVTKLDGTAKGGILAAIARQRPVPVPIGVGEQVEDLQPSAREFADALLG"
4124..5629
/gene="RSC0387"
/note="RSC03355"
4124..5629
/gene="RSC0387"
/EC_number="3.4.99.-"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
predicted by Homology
Predicted by FrameD"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE ZINC PROTEASE PROTEIN"
/protein_id="CAD13915.1"
/db_xref="GI:17427397"
/translation="WRSFQSVNSARTGLRWAGITLAVALLAGEAGAQAAPAPMARQ
AAALTGAKVTSQADTHERYLANGRLIVREDHAPTVAHVVWTHAGSIDHEHNTGV
AHMLHEHMFQTRAVGGEFSRRVAALGGRENAMTRDFTFYFOQIEKSLADVMAL
ADMANLTQTDKEFPENNVKEERRMRIDDSRATVVEQMLAVLFNAAPVNPRTFY
PSDLMTVQDAQKHVYAPNNAVTVVTDVNPDEVFOAORTYKGLQPHALPRY
AQDEPKVGKRIWKAAPENPVYLAAPPLDEKVDVDPVALEVLAVLGYDNA
FLPNLLVKGKDEKRAVDVAGYDNGSGSPSIFLIDGVPADGHTTABIETQALRAQI
DRIARDGTEALRKRAQVAAQIYKRDVSFGQMGMEIGMAEMTGLSWRDLRIEKI
KS2VPAQVQVAKTYFTEDNLVATLLPQPIDPNPKARKVPVPGNREEGGLR"
5629..6972
/gene="RSC0388"
/note="RSC03356"
5629..6972
/gene="RSC0388"
/EC_number="3.4.99.-"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon usage
predicted by Homology
Predicted by FrameD"
/codon_start=1

```

```

/evidence=not experimental
/transl_table=11
/product="PUTATIVE ZINC PROTEASE-LIKE SIGNAL PEPTIDE
PROTEIN"
/protein_id="CAD13916.1"
/db_xref="GI:17427398"
/translation="MRAMSLTLTKFLAGLMAAAHGAALALPIEHWTAAATGARVFFVP
SPSIPMLDINLDVADGTRYPAKVGSLASTAGMLDKGVAAGVSTPARDAAATADAF
DVGASPSGGAGGDRTSLSRLTSLDPAERKPAVDMQAQIVAAPTVPDAVLTRDQRTVA
Query Match 11.8%; Score 84.2; DB 1; Length 190050;
Best Local Similarity 47.9%; Pred. No. 3.4e-10;
Matches 242; Conservative 0; Mismatches 263; Indels 0; Gaps 0;
Q 207 CGTCTCGATATTATCGCTGGGTGGCATCGCTTACTTGTATGTTTCCGTCATGGC 266
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2423 CGTCTCGCGCTGTGTGATCTCGCGCTGTGCGACATGGCGCTGATCGCGCTG 2482
Q 267 AGCGAAAGACCCCATGACAAACAGGTGGAAGCGCCACAGATCATTTGAAGAAACAGAAC 326
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2483 GGCGGGCATGGCACGCTGATCTCGCGCACCCCGCATGCTGACGGGGTGGCGCTGGGC 2542
Q 327 AACCTGCCCCGATGACACGCTTTTGGCGGTTCGCGGTGGCCACTGACACGCGCAACCG 386
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2543 CGCGCGCGCTTCCTGCTGGCTACGCGCACGCGCGCTTCGCGCGCGCTGGCGCGCGC 2602
Q 387 GGTGGGTGGAGTGAGCTGATAGACAGCGGTTTGGTAAAGCCCATGTTTATGTCG 446
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2603 CGAGCGCTGCGAGCGCGCAACGCGCAAGCGCTCGCATGCGCAGGTGCTGGCGCTCGC 2662
Q 447 AATCGTCTGACCTGTTGAACCGCAATGCTATTTCGACGCGTTTGTGTTATTCGCGCG 506
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2663 GCTGCGCTGTGCTGCTCAATCGCAGCTATCTCGACACCGTGTGCTGCTGGCGCGC 2722
Q 507 GTCGCGCGCAATACCGCGCACACCGCGGTGATTTTCGCGCTGCGCGCTTCGCGCGC 566
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2723 CATCGCGCGCGCTACGCGATCGCGCAACGTCGCGCTTCGCGCGCGCGCATGTGCGC 2782
Q 567 AAGCTGATCTGTTCCGCTGGTGGCTTCGCGCGCAGCAGCATTTGTCACGCCCGCTGC 626
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2783 GTCGATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2842
Q 627 CAGCCCCAAGGTGTGCGCTGGATCAACGTCGTCGTCGCGCAGTTGTGATGACCGCATGGC 686
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2843 CAGCGCGCTGCGCTGGCGGTGCTCGATGCGCTGATCGCGCGGTGATGTGGCGCATGC 2902
Q 687 CATCAAACTGATGTGATGGTTAG 711
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D 2903 GCTGACGCTGCTGATGGCGGTAG 2927

RESULT 10
SCSF8 33285 bp DNA linear BCT 12-MAY-2002
LOCUS Streptomyces coelicolor cosmid 5F8.
DEFINITION AL357613 AL645882
ACCESSION AL357613.2 GI:20520783
VERSION
KEYWORDS amino acid decarboxylase; anti-sigma factor; integral membrane
protein; LuxR-family regulator; LysR-family transcriptional
regulator; membrane transport protein; oxidoreductase; regulatory
protein; RNA polymerase sigma factor; serine/threonine protein
kinase; stress-inducible protein; TetR-family transcriptional
regulator; thioredoxin reductase; threonine dehydratase;
two-component sensor histidine kinase.
Streptomyces coelicolor A3(2).
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 33285)
REFERENCE Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J.,
AUTHORS Kinashi, H. and Hopwood, D.A.
TITLE A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome

```



/note="SCSF8.05c, possible LuxR-family regulator, len: 988  
aa. Similar to several other putative regulators from  
Streptomyces coelicolor including: TR:Q9XAN4  
(EMBL:AL079355) SC4C6.06 (893 aa), fasta scores opt: 305  
z-score: 338.7 E(1): 7.7e-11 26.4% identity in 910 aa  
overlap and TR:Q92573 (EMBL:AL035569) SC8D9.18 (1091 aa),  
fasta scores opt: 857 z-score: 926.1 E(1): 0.36.5% identity  
in 1077 aa overlap. Contains a Prosite hit to P80017  
ATP/GTP-binding site motif A (P-loop) pfam match to entry  
P00196 GerE, Bacterial regulatory proteins, luxR family."  
/codon\_start=1  
/transl\_table=11  
/product="putative LuxR-family regulator."  
/protein\_id="CAB93733.1"  
/db\_xref="GI:8347028"  
/db\_xref="GSPREMBL:Q9K419"  
/translation="MGVGPOTSLPVGREDELARLTGVLERAGARAVLVAGDA  
GVGKRTLHEAGRAAAGATTVLGCHVDLGDVGLPYLFTTEILGVLAADRFPAVLA  
GHPVADRLUGGDDGTDAPSRLLPFGVAALITELADVAPLLVLEDLHWADQSR  
DLLRFLGRLQRPAGGAGHRLALFASRADDLHRRHPLRLAEILVRLPGVERLE  
LRPLPDSOVARLRLRRRLPDSVTVRIVERAEGNFAEELVAATDAPAGVPSGL  
ADVLIRTEQLSETAQOVLTAAGVRVGHDLRDVAGLPEEELSALRAVERQLL  
VSGDGAYSFHALAREAVADLLPGEARLHGAFARLLAGPDRSSAARAHYRE  
SHDLPALASLEAADHAQRCVAPAEELRHVERAULDLMTAVDAARAGPDAVTLTLR  
ASAAVAGHELHRAVSLTRSLAGLGQDADLEAARVRYTLAGNLLSVNLESAYYS  
SEALGLI PAEPSPSTVWAAATHVTAAQVGENETALRVARRALAEELAVTDARAD  
LLISLTLEGNSRTPGREGRLLEARELARAGNAPVELRALFNLATGCPSPGLEEC  
LPWASGDLARRSGLLSPPYREMYRLRLVQVTLGHWDVLESAEHAGERSAVDQ  
HALGGLVLAARDGSDVADARALLNGPDPWARMVAGVVLTDAAALRGDAEADRW  
MRSSVETITEAGTRPTVTLRLATLALSADVTITELRGDGDAGVARWSDTAAELLAD  
ARWAGHGDIMPQFEGQAWLAERAEWSVAGAAPDAARAVAAFTYQDTYERA  
RCRLDALLAADREAEATAGVREARLGLATVLRERVDLVRGRLAGPARAG  
AGAAVLTAEQDVLRLALGHSNRRIGBELFISAKTASVHVSNILAKLNAASRTEAVA  
VAHRQGLVAPEPTSSH"

Query Match 9.2%; Score 65.6; DB 1; Length 33285;  
Best Local Similarity 55.7%; Pred. No. 1.3e-05;  
Matches 147; Conservative 0; Mismatches 114; Indels 3; Gaps 1;  
QY 432 GCCCATGTTGATGGAATCGTCTGACCTGTTGACCCGATGCGTATTGACCGCTT 491  
Db 18542 GCCCGTCTCCTGCTGCGCGCTGACCTGCTCAACCGCACGCTTACCTCGACACCGT 18601  
QY 492 TGTGTTTATCGCGCGCGTCCGCGCGCAATACCGCGCACCGGACGATGATTTCGCGCGC 551  
Db 18602 GTTCTGCTGGCTCGTGGCGCGC---CGACCGGGCGCGCTGCGCTTGGCTT 18658  
QY 552 TCGCGGTTCCGCGGACGCTGATCTGTTTCCGCTGTTGGGTTTTCGCGCGACGACGATT 611  
Db 18659 CGGAGCGCGCGCGCGCGCTGCTGTTTCCGCGCGCTTTCGCGCGCGCGCTTACCT 18718  
QY 612 GTACGCGCGCTGTCAGCGCCCAAGTGTGGCGCTGGATCAACGTCGTCGTCGCGAGTTGT 671  
Db 18719 CGCGCGCTTCTGTTCGCGCGCGCTGCTGCGCGGCTTCTCGACGAGCTGTTGGCGCGCAC 18778  
QY 672 GATGACCGCATGGCCATCAAACT 695  
Db 18779 CATGATCGTCTCGCGCTTCTCCCT 18802

RESULT 11  
AE004852  
LOCUS  
DEFINITION  
Pseudomonas aeruginosa PA01, linear BCT 30-AUG-2000  
genome.  
ACCESSION  
AE004852 AE004091  
VERSION  
AE004852.1 GI:9950587  
KEYWORDS  
Pseudomonas aeruginosa.  
SOURCE  
Pseudomonas aeruginosa.  
ORGANISM  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
REFERENCE  
1 (bases 1 to 10542)

## AUTHORS

Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M.,  
Garber, R.L., Coulter, L., Tolentino, E., Westbrook-Wadman, S., Yuan, Y.,  
Brody, L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,  
Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.  
Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
opportunistic pathogen

JOURNAL Nature 406 (6799), 959-964 (2000)

MEDLINE 20437337

PUBMED 10984043

## REFERENCE

2 (bases 1 to 10542)

## AUTHORS

Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,  
Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,  
Lagrou, M., Garber, R.L., Coulter, L., Tolentino, E.,  
Westbrook-Wadman, S., Yuan, Y., Brody, L., Coulter, S.N.,  
Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,  
Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Salter, M.H.,  
Hancock, R.E.W., Lory, S. and Olson, M.V.

## Direct Submission

## TITLE

Submitted (16-MAY-2000) Department of Medicine and Genetics,  
University of Washington Genome Center, University Of Washington,  
Box 352145, Seattle, WA 98195, USA

## JOURNAL

## FEATURES

## source

## Location/Qualifiers

## I. 10542

## /organism="Pseudomonas aeruginosa"

## /strain="PA01"

## /db\_xref="taxon:287"

## complement (139..612)

## /gene="PA4360"

## complement (139..612)

## /gene="PA4360"

## /codon\_start=1

## /transl\_table=11

## /product="hypothetical protein"

## /protein\_id="AAG07748.1"

## /db\_xref="GI:9950588"

/translation="MRLOVKQLQKLSQBRQADVSNQFGGQKRLDOLASDLKAQO  
ESAQVLVAQLDGKLOTLAAOEKIKALQVELKTNKQALGADVAALKQGNPSQTI  
KGTDLLVLRQLDNRPAQGNATFADFRVQVNRNINLQAVQVQLQQLNTR"

## /gene="PA4361"

## complement (1529..2518)

## /gene="PA4361"

## complement (1529..2518)

## /codon\_start=1

## /transl\_table=11

## /product="probable oxidoreductase"

## /protein\_id="AAG07749.1"

## /db\_xref="GI:9950589"

/translation="MRILVTGATGFTGGRFARFALQGLSVRVSGRRADAVEHLVARG  
AEFVPGDLADPALVLRLCDVRAVHAGVWGPFRFLAANGLARSVVEACMRQ  
KYRRLVHLSPSIYFDGRDHLDLNEEYVPRPSDHYGATKYQAEQLVLSARDLGLVL  
ALRPFVVGAGDTSIFPRMIQAHKRLKILGNLRNVDFTSVHNLDAFSCLLAGE  
PALGVYINISNOPVFDVAVNYVMRQDLPPVGHLPYAVGVYGLAALNEGVCRLPG  
RPEVILFRIGMAVMAKNFTLDIRAREYLDYDPRVSLMTALDEFCAWMAAQGL"

## /gene="PA4362"

## 2703..3761

## /gene="PA4362"

## 2703..3761

## /gene="PA4362"

## /codon\_start=1

## /transl\_table=11

## /product="hypothetical protein"

## /protein\_id="AAG07750.1"

## /db\_xref="GI:9950590"

/translation="WASHMWIYGANGYTRLVAEQARQREGLTFLGGRNPAALHAIGS  
QIGLECRVFDLGDQACRQKQVVAHACGPFPSATSTPMIAACRAAGTHYVDITGE  
IAVFEQAHAGDAEAREAGIVCGVDFVPTDCLAACTKALPAQRALGPDGTSQ  
LSTGTAKTVEGLKFGCKTRENRLRDVPLGKRRIDIDFGRIKHAVTTPMGDVATAY  
YSTGPDIEVLPAPLLALGMRLLDPLPLGRQVQDWLKGQVQDKRIAGPDQARE  
RLRTVWGEARNRGERRTARLETANVYDITLHGVLAVRHLDDYQGGPGGYTPPSRL  
GACVESLPFGSRTIVIG"

## /gene="PA4363"

## complement (3762..4664)

## /gene="iciA"

## /note="PA4363"

## gene

## CDS

## gene

## CDS

## gene

## CDS

## gene

## CDS

## gene

## CDS



```
CDS      complement(3762..4664)
        /gene="icia"
        /codon_start=1
        /transl_table=11
        /product="inhibitor of chromosome initiation IciA"
        /protein_id="AAG07751.1"
        /db_xref="GI:9950591"
        /translation="MILFDYKLLAALAAVVEQGFERRAAQALGLSQSAVSQRIKLEA
RVGPVLVRETPPHPTDLGRLLNHNVOQLLEGDLQRWPNLDEGAPERLIALNA
DSLATWNAAGVDFCAERRVLLDIVEDQEVGLKRMAGEVAGCVGSGARPVAGARSL
LLGAMRVGLASDPDFIARHFRGVEAALAGVPAIVFGPDLLQHLKDLGVGGGFI
HHLCSSEGFVRLTAGLGMGLVPERQVQGLARGELVELLPQGVIDVPLYWYWRNG
GELLASITHELLARAGDLVRVSG"
        4762..5163
        /gene="PA4364"
        4762..5163
        /gene="PA4364"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAG07752.1"
        /db_xref="GI:9950592"
        /translation="MAGETSLRTLLRTLTPELNPQDYVFCFCAAGALPNGAEPLASFR
ERGLTVLERRQERLGLPYEYVAAMITLVHSSLTAVGLTAFAFALAEAGISCNV
MAGYFHDHLFVARDEGRRLAQLQRLAAEAH"
        5165..5767
        /gene="PA4365"
        5165..5767
        /gene="PA4365"
        /codon_start=1
        /transl_table=11
        /product="probable transporter"
        /protein_id="AAG07753.1"
        /db_xref="GI:9950593"
        /translation="MWQSYLNGILVAAGLIAIGAQNAFVLAQSLRREHLSVAALCV
FCDVAVLVGFGVGLKLLLENPTLAIARMGGIAFLTWYGLKALLRALRPDALGNRAE
TGPSSKAVGLAALAVLLNPHVYLDVTLVLIAGLGAQQAAPGAYALGAASASLWFEA
LALGAALWAPLARPATRWLLDLVAAWMLGMAQLLFRG"
        6114..6695
        /gene="sodB"
        /note="PA4366"
        6114..6695
        /gene="sodB"
        /codon_start=1
        /transl_table=11
        /product="superoxide dismutase"
        /protein_id="AAG07754.1"
        /db_xref="GI:9950594"
        /translation="WAPFLPLPYEKNALEPHISAETLEYHHDKHNTYVYVNLNLP
GTEPEGKSLRIVKSSGGIFNNAQVWNHTFYWNCLSPNGGGOPTGALADAINAARG
SFDKFEFTKTSVGTSGCGWGLVKKADGSLALASTIGAGNPLTSGDTPLLTCDWE
HAYIDYRNLRPKVYEAFWNLVWDFVAKPAA"
        6876..8939
        /gene="PA4367"
        6876..8939
        /gene="PA4367"
        /codon_start=1
        /transl_table=11
        /product="conserved hypothetical protein"
        /protein_id="AAG07755.1"
        /db_xref="GI:9950595"
        /translation="MKLDYRHSLSLKLRLVLLAALAVGVLSQAQIVFPAKAKQAV
SSDAQRIILAVRDPSTQAVSEDEMAMQVLEGLFOHEAVRQASIGHPGMLAEKR
PLLLPLFRWLTDPILQBERTPSIRLIGPPYSEYIGDKLTLDTPAGNFVITSEI
FISGLRLAAGLVLFVYHMLTKPLSKIEHLVSINPDRPSQHPQLKGHREL
GLWTTANQLLASIESHSLRREAENLLRISQYDFLTGNLQQLQQLDQILDGAG
RQRRVAVLCGLDDFKGINEQYTYQLGDQLLIALADRLRHSARLSGLRGDQPA
LVQADIPOPYEAAELASQILDLGAPEIDQHEVLRATIGITLFPEDGETKLLQK
AEQVTKLAKTSRNRYOFYIASVDSEMRRELEKDLRALQHEHLHLYVQPOVDYD
HRVGVGVALRWQHLHGFPVDPDLFPIAQCNGSIFSIGEWLDQACRQIREHWDQF
DRLRMVNLSTVQLHNLALPRVSNLQVRLKPARSLEVEVTEGLMEDISTAAQHLL
SLRAGALIAIDDFGTGYSSLSYLSKLSPLDKIKDKSFVQDLQDEODATVIRAIQL
KSLGMQVIAEGVETAQEAYIIAEGCNEGQGYLSKPLPARELTQYLRKQARRLSQAT
```

```
SSERP"
complement(8940..9842)
/gene="PA4388"
complement(8940..9842)
/gene="PA4368"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07756.1"
/db_xref="GI:9950596"
/translation="WDFDLWLFIAASGLCAWAGMQFVHRLARALLGDVPTSRIS
SPGFLVGLVEEGEPGLQAPLTGKPLWVRYRLEVEKESGREKAWRTVDKGAASE
AFSLRDATDACLVDPRGAEPVPLTOQRWEAFREAPIDTLRLLESFGTLVGGERLVRY
TEERLHAGEPLVALGEPRSSGGRGGLDAERAGAVIREWKGDFHCLLARPDSDNGE
LDERWNRVRLAARLEARDHRASSAPTRHRLGKPEAHPTFILSSQGGEDDLAVQRR
QAAGAALLCLGSAALLAGWLFSLRF"
complement(9844..10416)
/gene="PA4369"
complement(9844..10416)
/gene="PA4369"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG07757.1"
/db_xref="GI:9950597"
/translation="MSLTAIAFWVLLLAGYAVLVNGLVRLKHAVGKAWANIDVLL
KQRHDELPLKVEACKQYMOHERATLERVTAARNAVASARERADVNALGKAESGLRAGL
GQLFALAENYPOLKANESFQILSQISGLENGIADRRELYNEAVNLNNVRIEQFPDLL
IARVAGFKAAELLQFSEAEKADVDLKALFG"
BASE COUNT 1755 a 3724 c 3387 g 1676 t
ORIGIN

Query Match      8.9%; Score 63.4; DB 1; Length 10542;
Best Local Similarity 55.0%; Pred. No. 4e-05;
Matches 149; Conservative 0; Mismatches 116; Indels 6; Gaps 1;

QY 440  TGAATGGCAATCGTGCTGACCTGGTTGAACCGAATGGTATTTGGACGCGTTTGTGTTTA 499
      |||||
Db 5502  TGGCGGCACCTGGCGGTCACTGCTCAACCCCACTCTATCTGATACCGTACTTCTTA 5561

QY 500  TCGCGCGCGTGGCGGCAATACGGCGACACCGGACCGTGGATTTTCGCGCTGGCGCGT 559
      |||||
Db 5562  TCGGTTGCTCGCGCGCCACGAGG-----CGCGCGCGCGCTATGCTCGCGCGCG 5615

QY 560  TCGCGGCAAGCTGATCTGTTCCGCTGCTGGTTCGGCGGCGAGCAGATTGTCAGCC 619
      |||||
Db 5616  CCAGCGCTCGCTGATGTGTTCTTTCGCCCTCGCGCGCGGCGATGGCTGCGCCCT 5675

QY 620  CGCTGTCCAGCCCCCAAGGTGTGGCGCTGGATCAACGTCGTCGTGGCAGTTGTGATGACCG 679
      |||||
Db 5676  GCGTGGCGCGCGCGCCACCTGGCGCTGCTCGACCTGATGTTGGCGGCCCATGATGCTGG 5735

QY 680  CATTGGCCATCAAACTGATGTTGATGGGTTA 710
      |||||
Db 5736  GCATGGCGCGGCAACTGCTGTTCGCGGGATA 5766

RESULT 12
AE008841/c
LOCUS      AE008841 typhimurium L72, section 145 of 220 of the complete genome.
DEFINITION      AE008841 AE006468
VERSION      AE008841.1 GI:16421612
SOURCE      Salmonella typhimurium L72.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Salmonella.
REFERENCE      1 (bases 1 to 24336)
AUTHORS      McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
      Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
      Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
```

Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R. and Wilson, R. K.  
Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

JOURNAL Nature 413 (6858), 852-856 (2001)

MEDLINE 21534948

PUBMED 11677609

REFERENCE 2 (bases 1 to 24336)

AUTHORS The *Salmonella typhimurium* Genome Sequencing Project.

TITLE Direct Submision

JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of

Genetics, Washington University School of Medicine, 4444 Forest

Park Boulevard, St. Louis, MO 63108, USA

COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; [http://kinich.cifn.unam.mx:8850/db/regulondb\\_intro.frameset](http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset)

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

#### FEATURES

source  
1. .24336  
/organism="Salmonella typhimurium LT2"  
/strain="LT2; SGSC 1412; ATCC 700720"  
/db\_xref="ATCC:700720"  
/db\_xref="taxon:99287"  
/note="LT2"  
171. .733  
/gene="ygfA"  
/note="STM3061"  
171. .176  
/gene="ygfA"  
/note="putative RBS for ygfA; RegulonDB:STMS1H003104"  
185. .733  
/gene="ygfA"  
/note="similar to E. coli putative ligase (AAC75949.1); Blastp hit to AAC75949.1 (182 aa), 86% identity in aa 1 - 182"  
/codon\_start=1  
/transl\_table=11  
/product="putative ligase"  
/protein\_id="AAL21936.1"  
/db\_xref="GI:16421613"  
/translation="MIKRRALTPQQRFQGOAAARMLSFPVVAHTVAVFLSFD  
GELDTQPLSQLWRAGRVLPVLPFPSPGNLLFLHYTPQSVLVTNRKIQEPRLDVR  
DLVPLAKDLVLTPLVAFDDEGQRLGNGGGFYDRTLQNWQHKILPVGAHDCQLVVEK  
LPVKEWDIPLPAVTPSKIWEW"  
complement (1114. .2359)  
/gene="serA"  
/note="STM3062"  
complement (1114. .2346)  
/gene="serA"  
/EC number="1.1.1.95"  
/note="similar to E. coli D-3-phosphoglycerate  
dehydrogenase (AAC75950.1); Blastp hit to AAC75950.1 (410  
aa), 96% identity in aa 1 - 410"  
/codon\_start=1

/transl\_table=11  
/product="D-3-phosphoglycerate dehydrogenase"  
/protein\_id="AAL21937.1"  
/db\_xref="GI:16421614"  
/translation="MAKVSLEKDKIKFLIVEGVHQKALESLRAAGVTNIEYHKGALDA  
EQKASIRDAHFIFGLRSRTHLVEEVNAAEKVAICFCITGNQVDLNAARKGIYVF  
NAFPTSNTSVAEIVIGELLILLRGPENAKAHGVNKLAAAGSFARGKGLGIIGYG  
HIGTQGLIAESLGMHVFYDIENKPLGNATQVQHSLLANMSDVVSLHVPNASTK  
NMGAKEIALMKPSGLINAAGTGVVDIPALCDALATKHLAAGAIIDVFTEPATNSDP  
FTSPICEFONVILTPHIGSTQEAQENIGLEVAGKLIKYSNDGSLTSAVNFPEVSLPL  
HGGRRLWHIENRPGVITALNQIFAEQGVNIAAQYLQTSARMGYVIDIEADGDVAEK  
ALAMAKPIGTNRALLY"  
complement (2354. .2359)  
/gene="serA"  
/note="putative RBS for serA; RegulonDB:STMS1H003105"  
complement (2612. .3286)  
/gene="rpiA"  
/note="STM3063"  
complement (2612. .3271)  
/gene="rpiA"  
/EC number="5.3.1.6"  
/note="similar to E. coli ribosephosphate isomerase,  
constitutive (AAC75951.1); Blastp hit to AAC75951.1 (219  
aa), 98% identity in aa 1 - 219"  
/codon\_start=1  
/transl\_table=11  
/product="constitutive ribosephosphate isomerase"  
/protein\_id="AAL21938.1"  
/db\_xref="GI:16421615"  
/translation="MTQDELKKAAGVAAALQYVQPGTIVGVGTSTAAHFIDALGTWKG  
QIEGAVSSDASTTEKLGKIHVFDLNEVDSIGIYVDGADEINGHMGKGAALXG  
EKIASVASKFCIADASKQVDILGKFLPVEVIMPARSAVARQIVKLGRPEYRONV  
VTDNGNVILDVYGMELDIPALENAINAIPIGVTVGLFANRGADVALIGTDPGVKTV  
K"  
complement (3281. .3286)  
/gene="rpiA"  
/note="putative RBS for rpiA; RegulonDB:STMS1H003106"  
3407. .4318  
/gene="iciA"  
/note="STM3064"  
3407. .3412  
/gene="iciA"  
/note="putative RBS for iciA; RegulonDB:STMS1H003107"  
3425. .4318  
/gene="iciA"  
/note="lysR family; similar to E. coli replication  
initiation inhibitor, binds to 13-mers at oric  
(AAC75953.1); Blastp hit to AAC75953.1 (297 aa), 97%  
identity in aa 1 - 297; transcriptional regulator of dnaA  
and argK"  
/codon\_start=1  
/transl\_table=11  
/product="inhibitor of replication initiation"  
/protein\_id="AAL21939.1"  
/db\_xref="GI:16421616"  
/translation="MKRPDYRTLQALDAVIRGERPERAAKLCITQSAVSQRIKQLEN  
MGQPLLVTVTPRTEQGGKGLALLRQVELLEEWLGDEQTSGLLSLAVNADSL  
ATWLLPALAPVLADSPIRLNQVDETRQERLRGEVVGAVSIQHALPSCVLKDLG  
ALDYLFVASKPFAERYFPNGVRSLLKAPAFDHLDMHQAFQONFDLPFGVPC  
HTVNSEAFVQLARQCTTCMTPHLQIEKESELSGLINLTGQLQRRLMYWHRFADES  
RMWRKYTDALLEGYGHKVLQD"  
complement (4584. .5341)  
/gene="yggE"  
/note="STM3065"  
complement (4584. .5330)  
/gene="yggE"  
/note="similar to E. coli putative actin (AAC75959.1);  
Blastp hit to AAC75959.1 (246 aa), 88% identity in aa 1 -  
245"  
/codon\_start=1  
/transl\_table=11  
/product="putative periplasmic immunogenic protein"  
/protein\_id="AAL21940.1"

```
/db_xref="GI:16421617"
/translation="MKFKVMAALAVGLSAMSQAQSELPEGPPIHVTSGTASVDAVPDI
ATLAEVNAKDAATKQADERVAQYLPLEQNIAXKDIISAANLRTQPDYDQNG
KSILKGVRAVTVTRTQKLDKLSLGDGALKAGLINEIRSVLSGVAQPDYKQKARKA
AIDDAIHOAQAAGFNSKLPVSVRYHVSNYQPSVPMVMKAADAAPVSAQETVEQ
PTIQFDQDVVFOLEPCTGTSITTAATQ"
complement(5336..5341)
RBS
/gene="yggA"
/notes="putative RBS for yggB; RegulonDB:STWS1H003108"
complement(5424..6072)
gene
/gene="yggA"
/notes="yggA"
complement(5424..6059)
CDS
/gene="yggA"
/notes="similar to E. coli orf, hypothetical protein
(AAC75960.1); Blast hit to AAC75960.1 (211 aa), 90%
identity in aa 1 - 211"
/codon_start=1
/product="putative LYSE family amino acid transport
protein"
/db_xref="GI:16421618"
/translation="WISGVYFOGFAALGAMILPLGPONAFVNMQGIROYHLMIALICA
LSDVILSAGIFGSGSLLMOSPMILALVTWGVAFLLWYFGALKATAMSNLISLASAE
VMKGRKLIATMLAVTNPWHVYLDIFVVLGSLGGQLAMEPKRWFLPAGTSLFLWF
FGLALLAAWLAAPRURAKAQRINILVGVWMLIAFLQAREGVAMHAFN"
complement(6067..6072)
RBS
/gene="yggA"
/notes="putative RBS for yggA; RegulonDB:STWS1H003109"
complement(6259..7131)
gene
/gene="yggB"
/notes="yggB"
complement(6259..7119)
CDS
/gene="yggB"
/notes="involved in stability of Macs mechanosensitive
channel; similar to E. coli putative transport protein
(AAC75961.1); Blast hit to AAC75961.1 (286 aa), 91%
identity in aa 1 - 285"
/codon_start=1

Query Match      8.8%; Score 62.8; DB 1; Length 24336;
Best Local Similarity 51.8%; Pred. No. 6.4e-05;
Matches 142; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Qy 23 TTACAGGTCTGTTTGGGGCGAGCTTTTACTCTCATCGGACCGCAGAAATGACTGG 82
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6043 TTCAAGGGTTTGGCCCTGGCGCGCATGATCCTTCGCTTGGCCGCAAAATGCGTTG 5984

Qy 83 TGATTAAACAGGAATTAAGCGGAGGACGACTATTCGCTTCTTCTGTTGTTTAAATT 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5983 TGATGAATCAGGAATTCGTGCCAGTACCACCTCATGATACGCTGCTTTGCGCGTTAA 5924

Qy 143 CTGACGCTCTTTTGTTCATCGCGCGCACCTGGCGGTTGATCTTTTGCCAAATGCCGCGC 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5923 GTGATTTAGTCTGATTAGCGCGGATTTTGGCGGTAGCGGTTGCTGATGAGTGGC 5864

Qy 203 CGATCGTCTGATATTATGCGCTGGGTGGCATCGCTTACTCTGTTATGTTTGGCGTCA 262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5863 CGTGGCTCTGCGCTGCTGTCATCGGCGCGGTCGCTTCTTGTATGATGACGTTTGG 5804

Qy 263 TGGCAGCGAAGACGCCATGACAAACAGGTGGA 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5803 GCGCCTGAAACGCGCAATGAGCATATCTGGA 5770

RESULT 13
AE005522/c      10701 bp      DNA      linear      BCT 21-MAR-2001
LOCUS           Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 141
DEFINITION      of 290.
ACCESSION       AE005522 AE005174
VERSION         AE005522.1 GI:12517448
```

## KEYWORDS

SOURCE  
ORGANISMREFERENCE  
AUTHORSTITLE  
JOURNALMEDLINE  
PUBMEDREFERENCE  
AUTHORSTITLE  
JOURNALFEATURES  
Source

gene

CDS

gene

CDS

Escherichia coli O157:H7 EDL933.  
Escherichia coli O157:H7 EDL933  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia  
1 (bases 1 to 10701)  
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,  
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,  
Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,  
Grotbeck, E.J., Davis, N.W., Lim, A., Dimallanta, E., Potamousis, K.,  
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,  
Welch, R.A. and Blattner, F.R.  
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7  
Nature 409 (6819), 529-533 (2001)  
21074935  
11206551  
2 (bases 1 to 10701)  
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,  
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,  
Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,  
Grotbeck, E.J., Davis, N.W., Lim, A., Dimallanta, E., Potamousis, K.,  
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,  
Welch, R.A. and Blattner, F.R.  
Direct Submission  
Submitted (22-OCT-2000) Laboratory of Genetics, University of  
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
Location/Qualifiers  
1. .10701  
/organism="Escherichia coli O157:H7 EDL933"  
/strain="EDL933"  
/stratotype="O157:H7"  
/db\_xref="taxon:155864"  
/note="enterohaemorrhagic"  
133..2277  
/gene="sbm"  
/note="24254"  
133..2277  
/gene="sbm"  
/function="enzyme; Degradation of small molecules: Fatty  
acids"

/notes="Residues 1 to 714 of 714 are 99.15 pct identical to  
residues 1 to 714 of 714 from Escherichia coli K-12 Strain  
MG1655: B2917"  
/codon\_start=1  
/transl\_table=11  
/product="methylmalonyl-CoA mutase (MCM)"  
/protein\_id="AGS8043.1"  
/db\_xref="GI:12517449"  
/translation="MSNEQEWQOLANKLSREKTVDSLVOQTAEIGAIKPLYTEADL  
DNLVETGLPGLPPVYRGPRATYTAQPIITQIYAGFTAKESNATYRNLAAGQGL  
SVAFDLATHRGYDSNPRVAGDVGTGVAIDTVEDMKVLFDQIPLDKMSVMTMGAV  
LPVAFYITAEQGVTPDKLTQNDILKEYLCHNTYIYPPKPSMTIADIAWCS  
GNMPRENTISISGYHMGAGANCVQVATADGIEYIKAAISAGIKIDDFRLSFF  
FGIMDLFMNVAIRAYLVSEAVSGFGAQPDKSLARTHTCQTGWSLTEDQPPYNNV  
IRTTTALAATLGGTQSLTNADFALGLPTDFESAIABNTQIIIOESELCTVDPL  
AGSYTESLTDIVKQARAIIOQIDEGAKAKAEAGLPKRMIEEASAREQSLDQK  
RVTGVNKKYKLDHEDTDLVIDNVMVNEQIASLERIRATRDAAATTAALNATHAA  
QHNEILAAAVNAARVRATIGISDALEAFDRYLVPSQVTGVIQAQSYHQSSEKSAE  
FOATVAQTEQFLADNRPRRIIAKMGLDHGRKAKVISAAYSIDLGFVDLSPMFSTP  
EETARLAVENDVHVVGASSLAAGHTLIPELVEALKKWKREDICVVGAGVIPPQDYAF  
LOERGVAIYGGCTPMLDSVRDVLNLSQHHD"

2270..3265  
/gene="ygfD"  
/notes="24255"  
2270..3265  
/gene="ygfD"  
/function="phenotype; Not classified"  
/notes="Residues 1 to 331 of 331 are 99.69 pct identical to  
residues 1 to 331 of 331 from Escherichia coli K-12 Strain  
MG1655: B2918"  
/codon\_start=1  
/transl\_table=11  
/product="putative nucleotide-binding protein"

```

/protein_id="AAG58044.1"
/db_xref="GI:12517450"
/translation="MINEATLAESTIRLRQGEHATLAQMTLVESRPHRQALSTQLL
DAIMPCYCNLTGLVGTGPGAGKSTFLEAFGLLRGKLVAVIADVPSPVGTGSL
GKTRNDLRARAAIRPVPSSGHLGASQARREMLLCEAAGDYDVIVETVGVQS
ETEVARNDCFLQIAGGDDLLQKKGMEVADLIYNKDDGDNHTNVAIRHMYE
SALHTLRKYDEWPRLVTCSALEKRGIDEIWHAIIDFKTALTASGRLLQVQROQSV
WLRKQTEBEVLNHLFANEDFDYRYRQTLAVKNNTLSRPTGLRQLSEFIQTQYFD"
3234..4061
/gene="ygfG"
/note="y4256"
3234..4061
/gene="ygfG"
/function="putative enzyme; Not classified"
/note="Residues 1 to 275 of 275 are 100.00 pct identical
to residues 1 to 275 of 275 from Escherichia coli K-12
Strain MG1655: B2919"
/codon_start=1
/transl_table=11
/product="putative enzyme"
/protein_id="AAG58045.1"
/db_xref="GI:12517451"
/translation="NLSRRNLIKGFMSYQVNVVINKVAVIEFNYGRKLNALSK
VFIDMLQALSDLNPEIRCIILRAPSGKVSFAGHDIEHLPSSGRDPLSYDDPLRQI
TRMIQKFKPIISMVGGWGAFAEMSSDLIIAASSTFSMTGPNLGVNVLVGIH
NLTRDAGPHIVKELIFTASPIAORALAVGILNVVEVELEDFTLOWAHHISEKAPL
ALAVIKELRVLGAEHTWNSDEFERIQGMRRVYSEDYQSGMAFLKRPKNFVGH"
4085..5563
/gene="ygfH"
/note="y4257"
4085..5563
/gene="ygfH"
/function="putative enzyme; Not classified"
/note="Residues 1 to 492 of 492 are 99.39 pct identical to
residues 1 to 492 of 492 from Escherichia coli K-12 Strain
MG1655: B2920"
/codon_start=1
/transl_table=11
/product="putative coenzyme A transferase"
/protein_id="AAG58046.1"
/db_xref="GI:12517452"
/translation="METQWTRMTADEAAEIIQHNDMVAFSGFTPAGSPKALPTAIARR
ANEQHEAKKPQIRLLTGASISAAADDVLSDADAVSWRAPYQTSGLRKKINQAVSF
VDLHSEVAMQVNYGFDIDVAVIEASALAPDGRVLTSGIGNAPTLLRAKVIIE
LNHYDPRVLAELADIVIPAPPRNSVSIHAMDRVGYVQIDPKTIVAVETNLPD
AGMLDKQNPQCQIADNVVFLTKISASASSLTISADSLRKIYDNMDYPASRIVLRP
EIPPMWTSVLELVSVLLETGKISASASSLTISADSLRKIYDNMDYPASRIVLRP
QEISNPSIIRLGVIALNVGLEFDIYGHANSTHVAGVDLMNGIGSGDFERNAYLSI
FMASIAKEGKISTVPMCSHVHNEHSVKVITTEQGIADLRGLSPLOARTIINDCA
HPMYRDLRYLENAPEGHIIHDLSHVDFLHRNLITATSGMLG"
complement(5878..6456)
/gene="y4258"
complement(5878..6456)
/gene="y4258"
/function="putative regulator; Not classified"
/note="Residues 1 to 192 of 192 are 99.47 pct identical to
residues 6 to 197 of 303 from Escherichia coli K-12 Strain
MG1655: B2921"
/codon_start=1
/transl_table=11
/product="partial putative transcriptional regulator
LYSR-type"
/protein_id="AAG58047.1"
/db_xref="GI:12517453"
/translation="MDIFISKMRNFILLAQTNNAIAEAKIHTASPPFGKSIALAE
QIGYTLFRKDNKN="MDIFISKMRNFILLAQTNNAIAEAKIHTASPPFGKSIALAE
YPTIIFDOLISIGKEGVTAQPFVSENGVIDNLFDRQLDFIISPQHSARVQELN
LTISELPLRLNELVSRREYERQELLQLP"
complement(6623..7363)
/gene="yggB"
/note="y4259"
complement(6623..7363)
/gene="yggB"

```

```

/function="phenotype; Not classified"
/note="Residues 1 to 246 of 246 are 100.00 pct identical
to residues 1 to 246 of 246 from Escherichia coli K-12
Strain MG1655: B2922"
/codon_start=1
/transl_table=11
/product="putative actin"
/protein_id="AAG58048.1"
/db_xref="GI:12517454"
/translation="MKFKVIALAALMGISMAAQANELPDGPHIVTSGTASVDVDPDI
ATLAEIVNAAKDAAATKAKQADERVAQYISFLELQIAKDISSANLRTQPDYDQG
KSLIKGYRAVTVETVTLRLQDLKNSLLDGLAKLGNLEIRSVSLGVQAQPDAYKDKAKA
AIDNAIHQAQELANGFRLKGLPVYSRVYHVSQPSFPMVRMMKADAAPVSAQTYEQA
AIOFDQDVVFQLEFPVDOQPAKTPAAQ"
complement(7456..8091)
/gene="yggA"
/note="y4260"
complement(7456..8091)
/gene="yggA"
/function="orf; Unknown function"
/note="Residues 1 to 211 of 211 are 98.10 pct identical to
residues 1 to 211 of 211 from Escherichia coli K-12 Strain
MG1655: B2923"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG58049.1"
/db_xref="GI:12517455"
/translation="MFSYYPQGLAALGAMILPLGPQNAFVMNQGIROYHIMIALCA
ISDLVLCAGIFGSGALMQSPMLLALVTGWGVFLVLTGFGAFKTMAGSNIELASAE
VLKQGRWKIATMLAVTLNPHVYLDTFVVLGSLGOLDVXPKRWALGTISASFLWF
FGLAILAAMLAPELRITAKSQRIINLVVGCWMFIALQARDGIAHAQAQFS"
complement(8230..9090)
/gene="yggB"
/note="y4261"
complement(8230..9090)
/gene="yggB"
/function="putative transport; Not classified"
/note="Residues 1 to 286 of 286 are 100.00 pct identical
to residues 1 to 286 of 286 from Escherichia coli K-12
Strain MG1655: B2924"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AAG58050.1"
/db_xref="GI:12517456"
/translation="MEDLNVDISINGAGSLVANQALLSYAVNVAALAIITVGLII
ARMSINAVRLMISRKIDATVADFLSALVRYGIIAFTLIAALGRVGVQTASVAVLGA
AGLAVGLALQSLSLNLAAGVLLVMFRPFRAGEVVDLGGVAGTVLSVQIFSTWRTADG"
Query Match 8.7%; Score 62; DB 1; Length 10701;
Best Local Similarity 51.4%; Pred. No. 9.1e-05;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;
Qy 19 TTCATTACAGTCTGCTTTTGGGGGCCAGTCTTTTACTGTCCATCGCACCGCAATGTA 78
Db 8079 TACTTTCAAGTCTTCTCTTGGGGCAGCTATGATTCGCGCTCGGCCCCACAAAATGCT 8020
Qy 79 CTGCTGATTAAACAAGGAATTAAGCGCGAAGCAATCAATCGCGTTCTTCTCGTGTGTTA 138
Db 8019 TTTGTGATGAATCAGGCATTCGTGCTCAGTACCACATATGATTCCTTACTTTGGCT 7960
Qy 139 ATTTCTCAGTCTTTTGTTCATCGCGGCACCTTGGCGGTGTGATCTTTTGTCCAATGCC 198
Db 7959 ATCAGCGATTTGGTCTCTGATTTGGCGGGGATTTTTCGTGTAGCGGTATTATGATCAG 7900
Qy 199 GCGCGGATCGTCTGATATATGCGTGGGTGGCATCGCTACCTGTTATGTTTGGC 258
Db 7899 TCGCGGTTGTCTGGCGGTGTCACCTGGGGCGGCGTAGTCTTCTTGTGCTGTATGGT 7840
Qy 259 GTCATGCGCAGCAAGACGCGCATGACAAACAAGGTGA 296
Db 7839 TTTGGCGCTTTTAAACACAGCGATGATGATATATTTGA 7802

```



```

/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37077.1"
/db_xref="GI:13363126"
/translation="MSSYANQALAGLTGKSTDYRDYDASLLQGVPSRLNRPGLG
KADNLPFGDTIDWLYELSWLNAGLPQVAVGHVELDYTSVNLIESKSFKLYLNSFNO
TRFNWMDVROTLEKSTCAQGEVSVALYRLDELGQPIGHFNFGTCDDQDITDNY
EFTDYLBNATSGEKVVEETLVSHLLSNCLITQPDWGSIQIQRGRQIDREKLRY
LVSPRHNEFHEQCVERIFNDLLRQCPQKLSVARYTRRGGLDINPWSNSDFVPS
TRLVRQ"
4017..5381
/gene="EC83655"
CDS
4017..5381
/gene="EC83655"
/note="similar to YGDH ECOLI gi|1789159 percent identity
99 in 454 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB37078.1"
/db_xref="GI:13363127"
/translation="WITHISPLGSDMLSQLEVDMLKRTASSDIYQLFRNCSLAVLNS
CSLTDNSKELSRFENDIVRRRGVGLSLNPPPEAFVDGRIIRALQANLPAVIR
DILFYGGQHTNVRFPNLNDSVHTNLVFSILRNARFALHVGAPNNVCGGSHIN
ENBYLYARVGNQGLRELNTCTGCGPGEAPMKGAAGVHAQRYKDSRFIGMTSPS
ITAAEPNPVNELIIMPDIEKLEAFVRIAHGIIIPGGVGTAEELLYLGLIMNPA
NKDOVPLILGPKESADYRVLDFEVVHTLGENARRHYRIIDDAEVARQMKSNP
LVKNNRRDGDAYSFNMSBRTAPDQMPFEPESHENMANLKLPPQPEVLADLRAP
SGIVAGNVKEIGRIAEIEFGYKINGKEIMRRMDDLLQGFVAQHRMKLPGSAIYPC
EICT"
5938..7227
/gene="EC83656"
CDS
5938..7227
/gene="EC83656"
/note="similar to SDAC ECOLI gi|1789160 percent identity
100 in 429 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="probable serine transporter"
/protein_id="BAB37079.1"
/db_xref="GI:13363128"
/translation="METTQTSTIASKDSRSARWKTDTMMLGLYTAIGAGVFLPLIN
AGVGMPLIIMALPMTFAHRGLTRFVLGKNGPEDITVEVREHFGIGAKGLIT
LLYFPAIPLLYLVALTNTVESPMHQMTPTPPRAISLLILVGMWTVIRFGEOM
IVKMSILVFPFVGLMLALYLPQWGALETLSLDTASATNGNGLWTLWLAIPTM
VFSNHPSPITSSYANHFNAPVIAWMAPIIAITKSFGLHGARGGFGNVIK
DLAAKEQNTSILSYLANHFNAPVIAWMAPIIAITKSFGLHGARGGFGNVIK
SLRGKGSIEINKRITAPLMTVTWIVATNPSILGMETLGGPIIAMILLPMWY
AIQKVPAMKYSGHISNVFVVMGLIAISAFYSLS"
7285..8652
/gene="EC83657"
CDS
7285..8652
/gene="EC83657"
/note="similar to SDAA ECOLI gi|1788116 percent identity
77 in 455 aa, similar to SDAB ECOLI gi|1789161 percent
identity 99 in 455 aa (Conserved in E.coli K-12)"
Query Match      8.7%; Score 62; DB 1; Length 266658;
Best Local Similarity 51.4%; Pred. No. 0.00014;
Matches 143; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 19 TTCATTACAGGTCGCTTTGGGGCCAGTCCTTTTACTGTCCATCGGACCGCAGATGTA 78
Db 161581 TACTTCAAGGCTTGCTCTTGGGGCAGCATGATGTCGCGCTCGGCCCAAAATGCT 161522

QY 79 CTGGTGATTAAACAGGAATTAAGCGCGAAGGACTCATTTGGCGGTCTTCTCGTGTGTTA 138
Db 161521 TTTTGATGATCAGGCAATTCGTCGTAGTACCAATATGATTCCTTACTTTCGCT 161462

QY 139 ATTTCTGACGCTTTTGTTCATCCCGCGCACCTTGGCGGCTTGATCTTTTGTCCAAATGCC 198

```

```

Db 161461 ATCAGCGATTGGTCCTGATTTCGGCGGGGATTTTGGTGGTAGCGGCTTATTGATGCAG 161402
QY 199 GCGCGGATCGGCTCGATATTATCGCTGGGTGGGTCGATCGCTTACCTGTTATGTTTGGCC 258
Db 161401 TCGCGCGGTGTTGCTGGCGCTGGTCACTGGGCGCGCTAGTCTTCTTGGCTGTGATGGT 161342
QY 259 GTCATGGCAGCAAGACGCCATCACAAACAAAGGTGGA 296
Db 161341 TTTGGCGCTTTTAAACAGCGATGAGTAGTAATATGCA 161304

RESULT 15
ASU65741
LOCUS
DEFINITION
ASU65741 4033 bp DNA linear BCT 11-SEP-1997
Aeromonas salmonicida orfC replication inhibitor (icia), N-acyl
homoserine lactone synthase (asaI), transcriptional activator
(asaR) and YggA (yggA) genes, complete cds, and 2'3'-cyclic
phosphodiesterase (cdpB) gene, partial cds.
ACCESSION
U65741
VERSION
U65741.1 GI:1519234
KEYWORDS
Aeromonas salmonicida.
SOURCE
Aeromonas salmonicida
Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
Aeromonas.
REFERENCE
1 (bases 1 to 4033)
Swift, S., Karlyshev, A.V., Fish, L., Durant, E.L., Winson, M.K.,
Chhabra, S.R., Williams, P., Macintyre, S., and Stewart, G.S.
Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
identification of the LuxRI homologs AhyRI and AsaRI and their
cognate N-acylhomoserine lactone signal molecules
J. Bacteriol. 179 (17), 5271-5281 (1997)
JOURNAL
MEDLINE
97431471
PUBMED
9286976
REFERENCE
2 (bases 1 to 4033)
Swift, S., Karlyshev, A.V., Fish, L., Durant, E.L., Winson, M.K.,
Williams, P., Macintyre, S., and Stewart, G.S.A.B.
Direct Submission
TITLE
Submitted (01-AUG-1996) ABFS, University of Nottingham, Sutton
Bonington Campus, Leicestershire LE12 5RD, England
FEATURES
Location/Qualifiers
1..4033
/orIGIN="Aeromonas salmonicida"
/strain="NCIMB 1102"
/db_xref="taxon:645"
/complement (30..929)
/gene="icia"
/complement (30..929)
/notes="icia"
/codon_start=1
/transl_table=11
/product="orfC replication inhibitor"
/protein_id="AAB70016.1"
/db_xref="GI:1519235"
/translation="MKALDYKLLALLADVMQSFERRAQRHLITQSALISORIKOLEQ
QPAEPLIIRSQPLQATPLQKLLAHYRQVRLQELAGIADDEQAPRVSIVNAD
SLATWFLPALAPLELQHPLELNLVDDECTLDVRQGFAGVSLHGQPLAGCCVDE
LGEMRYLLTASPAFVAHPFAGLTPAALAKTFAVAFDQDDMHVSMFARHFGLEPGY
PCHTRVSSFAFVMAEQGLAYCLTPELQIRQOLAQGLILLDLSPSHLLIPLYWHRWVL
ERGLHKQISQRLISGRALQPG"
/complement (1195..1818)
/gene="asaI"
/complement (1195..1818)
/notes="asaI"
/codon_start=1
/transl_table=11
/product="N-acyl homoserine lactone synthase"
/protein_id="AAB70017.1"
/db_xref="GI:1519236"
/translation="MLVFKGLKHEPRMEVENELYRFRNRFSDRLGMDVESHGRLQ
DSFDTPDTHWVLIIEDEGLCGICLLSCAQDYMPLSIFPTALAGEAPRSSDVWELTR

```

Job time : 2268.29 secs

LAI DAN RAP RM GNG VSE L T CV I FRE VY AFA RAK G IRE L VAV VSL PVER I FRRL GLPIE  
RLG HRQAV D LGAVRG VGI RFH LDER FARA VGH PMOGEYADARELVTE "

gene  
1881, .2663

```
/gene="asaR"
```

```
1881. .2663
/ gene="aaaR"
/ note="AaaR: similar to Photobacterium fischeri LuxR"
/ codon_start=1
/ trans1_table=1
/ product="transcriptional activator"
/ protein_id="AAE70018.1"
/ db_xref="GI:1519237"
```

```

/translation="MKQDQLLEYLHFTSVTDGDRLAELIGPTLGMGYDYVRPALII
PMSGQKVLVFNQCDPSWQYFAATHNMLACDPI IQARKQTLPIVYNNRLDERARFLQ
EGSLDVMGLAELGRLNGIYPTLHGAAGACDPI IATRAERASDLLESPSLWSMNS
YIEFAAIRIVRSURDEQDQALTRDTETCLFWASEGKTSIGACILGITERVNYHL
NQVTRKTSGNRYQIA IAKGVSSTGILLPNLEWQVYVTFNPKLMO"

```

gene 2908. .3486

```
/gene="yggA"
```

```
2908..3486
/gene="yggA"
/notice="similar to Escherichia coli YggA"
/codon_start=1
/transl_table=11
/product="YggA"
/protein_id="AB070019.1"
/db_xref="GI:1519238"
```

```
/translation="MIPIGAQNAFVLSGRIHRNHLLAATJCCLCDLLIGIGVFGG  
ANLLAASPVGLIALTWGGVFLCWFGFIRSLRSANWQQAALADSPLRMGVKSVLANTL  
GVTLNAPHVGLDITLMLGSLFGSQFAEPIRPAPFAGMGLASLVWFYSILAFIGAALSPWL  
GRTLVQQAIDITVIGLMLGLAQALSGALLAS"
```

gene 3614. .4033

```
/gene="cdpB"
```

```
3614. .>4033
/gene="cdpB"
/note="CdpB; similar to Escherichia coli CdpB"
/codon_start=1
/transl_table=11
```

```

/product="2,3'-cyclic phosphodiesterase"
/protein_id="AAB70020.1"
/db_xref="GI:1519239"
/tranlation="MKLGAIAlVLLSACSNNDSKTPGAVARLLQTSDIHNVILG
YDYQNPKEKFGSLSTLIRAESENPNLLDNGLLQGTPLADYIPEQSGAGYL
DKQAPYFKEKMEIYVDAGNI GNHRNPGYLDI,TKTTL"

```

787 a 1160 c 1229 a 857 f

## ORIGIN

Query Match 8.7%; Score 61.6; DB 1; Length 4033;  
Best Local Similarity 52.3%; Pred. No. 0.0001;  
Matches 136; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 437 TGTGATGGCAATCGTGTGCTGACCTGGTTGAACCCGAATGCGTATTTGGACGCGTTTGTGT 496

Db 3200 TGGCCATGACGCTCGGGGTGACCTTGCTCAATCCTCACTATCTGGATACCCCTGATGC 3259

Qy 497 TTATCGGCGGCTCGCGCGCAATACGGCGACACCGACGTGGATTTCGCGCGTGGCG 556

Db 3260 TGCTGGGCTCTTTGGCAGCCAGTTTGCAGAA<sup>CGTTGGCCCGCGCCCTTTGGCGGCCGGGG</sup> 3319

Qy 557 CGTTCGGCGCAAGCCCTGATCTGGTTCCTGGTGGTTCGGCGCAGCAGCATTTGCAC 616

Db 3320 CCATGCTGGCTTCCCTGGTCTGGTTCTACAGTCTGGCCTTTGGTGGCGCGGCTTCAC 3379

QY 617 GCCCGCTGTCCAGCCCCCAAGGTGTGGCGCTGGATCAACGTGCTGCTGGCAGTTGTGATGA 676

Db 3380 CCTGCTTGACAGGGGCGGGTGAGCAAGCTATTGTACTACTATTTGTTGGTTTAAATCATGT 3439

677 CCGCATTGGCCATCAAACTG 696

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

Search completed: April 26, 2003, 23:30:20



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:02:03 ; Search time 947.401 Seconds  
(without alignments)  
12154.292 Million cell updates/sec

Title: US-09-105-117k-1\_COPY\_1016\_1726

Perfect score: 711

Sequence: 1 atgtgtatcatggaatctt.....aactgatgttggtggtag 711

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	60.4	8.5	830	17	CNS01MQH
C 2	48.6	6.8	925	17	CNS0091P
C 3	44.4	6.2	644	14	BQ704105
C 4	42.8	6.0	532	10	AW227407
C 5	41.2	5.8	910	17	CNS0060N
C 6	40.2	5.7	613	10	BE498441
					AL151258 Anopheles
					AL053013 Drosophil
					BQ704105 946110C04
					AW227407 945001E08
					AL065629 Drosophil
					BE498441 WHE0967_G

C 7	40	5.6	925	17	CNS0091P
C 8	39.6	5.6	1667	11	AV108197
C 9	39	5.5	950	11	AV104298
C 10	38.8	5.5	557	10	AW787854
C 11	38	5.3	903	13	BG975548
C 12	37.8	5.3	804	17	CNS01HIJ
C 13	37.6	5.3	370	13	BJ295166
C 14	37.2	5.2	451	17	BH843100
C 15	37.2	5.2	644	17	CNS0ED94
C 16	37.2	5.2	2084	12	BF128512
C 17	37	5.2	487	12	BF146009
C 18	37	5.2	677	13	BJ214340
C 19	37	5.2	911	17	CNS031CW
C 20	36.8	5.2	482	12	BG048617
C 21	36.8	5.2	506	13	BI683498
C 22	36.8	5.2	552	10	AW659102
C 23	36.8	5.2	562	17	AZ210936
C 24	36.8	5.2	623	13	BJ482071
C 25	36.8	5.2	677	13	BJ482072
C 26	36.8	5.2	692	17	CNS007WH
C 27	36.8	5.2	696	9	AA952035
C 28	36.8	5.2	843	17	CNS011A6
C 29	36.8	5.2	920	13	BI948315
C 30	36.6	5.1	583	13	BI099110
C 31	36.6	5.1	658	17	BH114920
C 32	36.6	5.1	839	17	CNS004NB
C 33	36.4	5.1	660	13	BI328066
C 34	36.4	5.1	827	12	BF627146
C 35	36.4	5.1	1110	12	BE788174
C 36	36.2	5.1	656	13	BI306486
C 37	36.2	5.1	682	14	BQ806138
C 38	36.2	5.1	986	17	CNS040T1
C 39	36.2	5.1	1011	17	CNS0209L
C 40	36	5.1	387	10	AV932366
C 41	36	5.1	689	13	BI248591
C 42	35.8	5.0	935	17	CNS006XK
C 43	35.6	5.0	166	10	AW351850
C 44	35.6	5.0	222	12	BF352878
C 45	35.6	5.0	404	13	EM068559

## ALIGNMENTS

RESULT 1  
CNS01MQH/c

LOCUS  
DEFINITION

CNS01MQH 830 bp DNA linear GSS 14-JUN-2001  
Anopheles gambiae GSS T7 end of clone 22E24 of NotreDamel library  
from strain PEST of Anopheles gambiae (African malaria mosquito),  
genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

REFERENCE

AUTHORS

TITLE

```

FEATURES
source
Location/Qualifiers
1. .830
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="22E24"
/clone_lib="NotreDamel"
/notes="end : T7"
BASE COUNT 207 a 231 c 227 g 156 t 9 others
ORIGIN
Query Match 8.5%; Score 60.4; DB 17; Length 830;
Best Local Similarity 50.7%; Pred. No. 5.4e-07;
Matches 142; Conservative 1; Mismatches 137; Indels 0; Gaps 0;
QY 23 TTACAGTCTGCTTTTGGGGCCAGTCTTTTACTGTCCATCGGACCGCAGAAATGACTGG 82
Db 597 KTCAGGGCTGCTTAGGTGCAATTTTATCTTCCCTCGGCCGCAAAATGCGTTGCG 538
QY 83 TGATTAACAAGGAATTAAGCGGAAGACATCATGCGGTTCTTCTGCTGTTTAATTT 142
Db 537 TGATGAATCAGGGCATCGCGCCAGTACCATCTGATGATTTGCCCTGCTATGCGCGGTA 478
QY 143 CTGAGCTTTTGTTCATCGCGGCACCTTGGCGGTTGATCTTTTGTCCAATGCCCGC 202
Db 477 GTGATTTGCTGCTGATCTGTGCGGGATTTTGGCGGACGCGCTGCTGATGCGATCTC 418
QY 203 CGATCGTCTCGATATATGCGTGGGTGGCATGCTTACCTGTATGTTTGGCGTCA 262
Db 417 CGTGGCTGCTGGCTTGTCTACCTTGGGCGCGTGGCGTTCTGCTCTGTGACGATCG 358
QY 263 TGGCAGCGAAAGACGCCATGACAAACAGGTGGAAGCGCC 302
Db 357 GCGCGCTGAAACGCGATGAGCAGTAACCTGCTGAACTGGC 318

RESULT 2
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammos in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
Location/Qualifiers
1....925
source

```

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/notes="end : TET3"
BASE COUNT 120 a 61 c 61 g 172 t 511 others
ORIGIN
Query Match 6.8%; Score 48.6; DB 17; Length 925;
Best Local Similarity 13.8%; Pred. No. 0.0021;
Matches 49; Conservative 157; Mismatches 149; Indels 0; Gaps 0;
QY 325 CCAACCGTCCCGATCACACGCTTTGGCGGTTGCGGCTGCGCCACTCACACGCGCAAC 384
Db 571 SCSCSSSCSSSCSCCCHCCSCSCSYCCSSSSSSSSSSSSSTSSBBSCTSSSSSS 630
QY 385 CGGTCGCGGTGAGCTGAGCGTTCGATAGCAGCGGTTTGGGTAAAGCCCATGTTGATG 444
Db 631 CSSSSSSSSSSTSSSTSSSTSSSSSSSSSSSSSSSTSSBBSCTSSSSSS 690
QY 445 GCAATCGTCTGACCTGGTTGAACCGAATGCGTATTGACGCGTTGTTGTTATCGGC 504
Db 691 STGTSSTSSSSSTSSSSSVSSGSKSTBSSGSSSSSSSSSTSSBBSCTSSSSSS 750
QY 505 GGGTCGCGCGCATACGCGACACCGACGCGTGGATTTTCCGCGCTGCGCGTTGCGG 564
Db 751 SSYSSTCTCCCTCCSYSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSS 810
QY 565 GCAAGCTCATCTGTTCCCGTGGTGGTTCGGCGCAGCAGCATTTGTCACGCCCGCTG 624
Db 811 MTCCCTYBMCVTSTSCGSSSSSSSGKGVTKGCGCGSSSTNGMBGTSSACSSSSSS 870
QY 625 TCCAGCCCAAGTGTGGCGCTGATCAACGCTGTCGTGCGCAGTTGTGATGACCG 679
Db 871 SSSSVSSSKSSASSSSSVSSGSSVSSNSNSASKSSSSGSSVSSGSSGSSGSSVSG 925

RESULT 3
BQ704105 644 bp mRNA linear EST 16-JUL-2002
LOCUS 946110C04.y1 946 - tassal primordium prepared by Schmidt lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION BQ704105
VERSION BQ704105.1 GI:21843524
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC
ciade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 644)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2327
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 946110 row: C column: 04.
Location/Qualifiers
1. .644
/organism="Zea mays"
/cultivar="OH43"
/db_xref="taxon:4577"
/clone_lib="946 - tassal primordium prepared by Schmidt
lab"
/tissue type="tassels"
/dev stage="just after the transition from vegetative to
inflorescence development"

```

/lab\_host="XL0LR"  
 /note="Organ: tassels; Vector: HybridZAP; Site 1: EcoRI;  
 Site 2: XhoI; George Chuck dissected immature tassels  
 between 1mm and 3mm. Sharon Stanfield prepared the cDNA  
 library in HybridZAP. Sample insert size range was 350 bp  
 to 3 Kb with a 1 Kb average."  
 BASE COUNT 118 a 239 c 180 g 107 t  
 ORIGIN

Query Match 6.2%; Score 44.4; DB 14; Length 644;  
 Best Local Similarity 46.0%; Pred. No. 0.033;  
 Matches 150; Conservative 0; Mismatches 176; Indels 0; Gaps 0;  
 QY 349 TTGGCGGTTTCGGGTCGACACGCGCAACCGGGTTCGGGGTGGAGGTGAGCGTC 408  
 |||||  
 Db 364 TTGGCGGTCGACGCGGAGAGTTCGAGATGCGCCCTTGGAGAGCTTCCGGTTATGCTC 305  
 |||||  
 QY 409 GATAGACGCGGTTTGGGTAAGCCCATTTGATGGCAATCGTCTGACCTGTTGAAC 468  
 |||||  
 Db 304 CTCCTGTAGACGAGGTGGAGCGGAGTGGACGTAGCAGGTGCCGGCGAGGTGGATC 245  
 |||||  
 QY 469 CCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGCTCGCGCGCAATACGGCGAC 528  
 |||||  
 Db 244 TCGAAGCGCCGTCGCTCTGAGCTGTAGAGGCGAGGAGTCCGGATGAGCCCC 185  
 |||||  
 QY 529 ACCGACGCGTGGATTTTCGCCGCTGGCGCTTCGCGGCAAGCCTGATCTGTTCCCGCTG 588  
 |||||  
 Db 184 TTCGGGAGCGGCTACTTGGGAGTAGTTCGTCGCGCGCTGAGCGAGCTGGCTGAC 125  
 |||||  
 QY 589 GTGGGTTTCGCGGACGACGATTTGACGCGCGCTGTCCAGCCCAAGGTGTGGCGTGG 648  
 |||||  
 Db 124 GCCGTGACCGCTGCGCGCGCGCGCGCGGACGAGGACGATCGCGGGCGAGCGGA 65  
 |||||  
 QY 649 ATCAACGTCGCTGGCAGTTGTGAT 674  
 |||||  
 Db 64 GACATTGCCGTTCGGATATTGGGAT 39  
 |||||

RESULT 4  
 AW927407/c  
 LOCUS  
 DEFINITION 945001E08.Y1 945 - Mixed adult tissues from Walbot lab, same as 707  
 (SK) Zea mays cDNA, mRNA sequence.  
 ACCESSION  
 VERSION AW927407  
 KEYWORDS EST.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.

REFERENCE  
 AUTHORS Walbot V.  
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Walbot V  
 Department of Biological Sciences  
 Stanford University  
 855 California Ave, Palo Alto, CA 94304, USA  
 Tel: 650 723 2227  
 Fax: 650 725 8221  
 Email: walbot@stanford.edu  
 Plate: 945001 row: E column: 08.  
 Location/Qualifiers  
 1. 532  
 /organism="Zea mays"  
 /cultivar="W23"  
 /db xref="taxon:4577"  
 /clone\_lib="945 - Mixed adult tissues from Walbot lab,  
 same as 707 (SK)"  
 /tissue\_type="tassel, kernal, silk, husk, root, leaf"  
 /dev\_stage="fully-grown"

FEATURES  
 source

/lab\_host="DH10B"  
 /note="Organ: tassels, kernal, silk, husk, root, leaf;  
 Vector: pGAD10; Site 1: EcoRI; cDNA library from fully  
 differentiated maize tissues from an active Mutator plant.  
 Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,  
 root, leaf). Unidirectionally cloned. New library number  
 given to library 707 for additional sequencing."  
 BASE COUNT 85 a 206 c 152 g 89 t  
 ORIGIN

Query Match 6.0%; Score 42.8; DB 10; Length 532;  
 Best Local Similarity 45.7%; Pred. No. 0.095;  
 Matches 149; Conservative 0; Mismatches 177; Indels 0; Gaps 0;  
 QY 349 TTGGCGGTTTCGGGTCGACACGCGCAACCGGGTTCGGGGTGGAGGTGAGCGTC 408  
 |||||  
 Db 332 TTGGCTGGACGCGGAGAGTCCGAGATGGCGCCCTTGGAGAGCTTCCGGTTATGGTC 273  
 |||||  
 QY 409 GATAGACGCGGTTTGGGTAAAGCCCATTTGATGGCAATCGTCTGACCTGTTGAAC 468  
 |||||  
 Db 272 CTCCTGTAGTACGAGGTGGGAGCGGAGTGGACGTAGCAGGTCCGGGAGTGGATC 213  
 |||||  
 QY 469 CCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGCTTCGCGCGCAATACGGCGAC 528  
 |||||  
 Db 212 TCGAAGCGCGCCCTCGTCTGAAAGCTGTAGAGGCGGACGAGTCCGGATGAGCCCC 153  
 |||||  
 QY 529 ACCGACGCGTGGATTTTCGCCGCTGGCGCTTCGCGGCAAGCCTGATCTGTTCCCGCTG 588  
 |||||  
 Db 152 TTCGGGAGCGGCTACTTGGGAGTAGTTCGTCGCGCGCTGAGCGAGCTGGGTGAC 93  
 |||||  
 QY 589 GTGGGTTTCGGCGCAGCAGCATTTGTCACGCGCGCTCTCCAGCCCAAGGTGTGGCGCTGG 648  
 |||||  
 Db 92 GCCGTGACCGCTCGCGCGCGCGCGCGGACGAGGACGATCGCGGGCGAGCGGA 33  
 |||||  
 QY 649 ATCAACGTCGCTGGCAGTTGTGAT 674  
 |||||  
 Db 32 GACATTGCCGTTCGGATATTGGGAT 7  
 |||||

RESULT 5  
 CNS0060N  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #  
 BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit  
 fly), genomic survey sequence.  
 ACCESSION AL065629  
 VERSION AL065629.1 GI:4944698  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 Ephydroidea; Drosophilidae; Drosophila.  
 Genoscope.  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
 - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a  
 collaboration with the Berkeley Drosophila Genome Project (BDGP).  
 The BDGP is constructing a physical map of the Drosophila  
 melanogaster genome using these BACs. For further information  
 please see <http://www.fruitfly.org/TheBDGP>  
 Aaron Mammoser in Pieter de Jong's laboratory in the Department of  
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
 NY. The library is named RPCI-98 and was constructed by partial  
 EcoRI digestion of Drosophila DNA provided by the BDGP from  
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's  
 P1 and EST libraries. A more detailed description of the library  
 and how to order individual BAC clones, the entire library, or  
 filters for hybridization from the BACPAC Resource Center can be

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

FEATURES
  source
    1. .910
      Location/Qualifiers
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="BACR14J21"
        /clone_lib="RpCi-98"
        /notes="end : T7"
BASE COUNT      202 a      63 c      112 g      198 t      335 others
ORIGIN
Query Match      5.8%; Score 41.2; DB 17; Length 910;
Best Local Similarity 14.8%; Pred. No. 0.35;
Matches 47; Conservative 150; Mismatches 119; Indels 2; Gaps 1;
QY 333 GCCGATACACGCTTTGGCGGTTGGCGGTCGACACGCGCAACCGGTCG 392
Db 518 GSSSKGGCGSGYGGGGYGGTGGSGGGSSSSSSSSSSSSSSSSSSSSSSSSB 577
QY 393 GGTGAGTGAGCTGCATAGACAGCGGTTGGTAAAGCCCATGTTGATGGCAATCGT 452
Db 578 SSCGCGSCYSSSTGTGCGCGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 637
QY 453 GCTGACC--TGGTTGAACCGAATGCGTATTGGACGCGTTTGTATTATCGCGCGTC 510
Db 638 SSTBTSSSTTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBT 697
QY 511 GCGCGCAATACGCGACACGCGGTTGCGCGGTCGCGGTTTCGCGGCAAGC 570
Db 698 SSSSTSSSTSTKBYSTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBT 757
QY 571 CTGATCTGTTCCGCTGGTGGTGGCGCAGCAGCATTTGTCACCGCGCTGTCAGC 630
Db 758 YTTBBSSTKBTSTBTSTKBTSTBTSTBTSTBTSTBTSTBTSTBTSTBTSTBT 817
QY 631 CCCAAGGTGCGCGCTGG 648
Db 818 SBCTSTSTSSBSBSBS 835

RESULT 6
LOCUS      BE498441
DEFINITION BE498441 613 bp mRNA linear EST 04-AUG-2000
            wheat pre-anthesis spike cDNA library Triticum
            aestivum cDNA clone WHE0967_G03_M05, mRNA sequence.
ACCESSION  BE498441
VERSION     BE498441.1 GI:9697058
KEYWORDS    EST.
SOURCE      bread wheat.
ORGANISM    Triticum aestivum
REFERENCE   1 (bases 1 to 613)
AUTHORS     Anderson,O.D.,Chao,S.,Choi,D.W.,Close,T.J.,Fenton,R.D.,Han
            ,P.S.,Heia,C.C.,Kang,Y.,Lazo,G.R.,Miller,R.,Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
TITLE       The structure and function of the expressed portion of the wheat
            genomes - Pre-anthesis spike cDNA library
JOURNAL     Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oanderson@wv.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
            Location/Qualifiers
              1. .613
                /organism="Triticum aestivum"

```

```

/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE0967_G03_M05"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EORI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give pluscript
phagemids in the TJ Close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT      144 a      178 c      188 g      102 t      1 others
ORIGIN
Query Match      5.7%; Score 40.2; DB 10; Length 613;
Best Local Similarity 53.5%; Pred. No. 0.61;
Matches 84; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 380 GCACCGGTCGGGTGAGGTGAGCTGCATAGCAGCGGTTGGTAAAGCCCATGT 439
Db 84 GCGAAGGATGCGCGCGGAGGGGAGCGCATCCGCTGTTCCGCGTCCGCGCACGG 143
QY 440 TGATGCAATCGTGTGCTGACCTGTTGAACCCGAATCGTATTGGACGCGTTGTTGTTTA 499
Db 144 TGATGCAATCGTGTGCGGACCGCGGTACTGTTAGACCGCGACGTGCAGTACCA 203
QY 500 TCGCGCGCTCGCGCGCAATACGCGACACCGGACG 536
Db 204 TGGCGAGTTGTCGCGGCGCTACGCGACCCCGTGGC 240

RESULT 7
LOCUS      CNS0091P/c
DEFINITION CNS0091P 925 bp DNA linear GSS 03-JUN-1999
            Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RpCi-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL053013
VERSION     AL053013.1 GI:4934461
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
ORGANISM    Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 925)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
            - Web : www.genoscope.cns.fr
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RpCi-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
            F1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
            Location/Qualifiers

```

```
source
1. .925
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT      120 a      61 c      61 g      172 t      511 others
ORIGIN

Query Match      5.6%; Score 40; DB 17; Length 925;
Best Local Similarity 13.7%; Pred. No. 0.81;
Matches 51; Conservative 157; Mismatches 164; Indels 0; Gaps 0;

QY 197 CCGGCCGATCGTCTCGATATTATGCTGGGTGGCATCTTACCTGTATGCTTTG 256
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 925 CSBSCSCSCSBSBSSSMTSSNSBSCSSBSSSSTSSMSSBSSSSSSSS 866

QY 257 CCGTCATGGCAGCAAGACCCATGACAAACAGGTGGAGCCGACATATTAAG 316
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 865 SSGTSSACVKCNASSCCGCCGABCCMCSSSSSCGASARGVKVRASGAGKRGGS 806

QY 317 AAACAGAACCAACGCTGCCATGACACGCTTTGGGGGCTTCGCGGTGCCACTGACA 376
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 805 GGASASHSSSACBSSSSCSASCMWSASSSSASRSRSGGAGGSCASSRSSSSSS 746

QY 377 CGCGAACCGGGTGGGTGAGGTGACGCTCGATAAGCAGCGGTTTGGTAAAGCCCA 436
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 745 ASAGSVSSASSSSSSSSSVSCSVASSMCSBSSSSASASSSSSSSSASCSCC 686

QY 437 TGTTCATGGCAATCGTCTGACCTGTTGAACCGAATCGTATTGACGCGTTGTGT 496
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 685 TSWSCSTASMAAARSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSGACG 626

QY 497 TTATCGCGCGCTCGCGCGCAATACGCGACACCGAGCGTGGATTTCGCGCTGGCG 556
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 625 SMSSGSGSVSAGSMSSSVSSSGRSSGGGGVGGSSGSGSGSGSGSVCS 566

QY 557 CGTTCGCGCAA 568
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 565 SGCMCRSCSSA 554

RESULT 8
AY108197
LOCUS      AY108197      1667 bp      mRNA      linear      HTC 25-MAY-2002
DEFINITION Zea mays PC0065133 mRNA sequence.
ACCESSION  AY108197
VERSION     AY108197.1 GI:21211275
KEYWORDS   HTC.
SOURCE      Zea mays.
ORGANISM   Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 1667)
            Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S.,
            Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
            Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
            Unpublished (2002)
            2 (bases 1 to 1667)
            Coe, E.C.
            Direct Submission
            Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
            Location/Qualifiers
            1. .1667
            /organism="Zea mays"
            /db_xref="MaizeDB:633502"
            /db_xref="taxon:4577"
            /clone="PC0065133"
            /clone_lib="Maize Mapping Project/DuPont Consensus
            Library"
            /note="this sequence is part of a project of EST
            assemblies resulting from the application of public
            contigs to seed DuPont contigs; this resource was
            assembled by DuPont as part of a collaboration for the
            overgo addressing of BACs in conjunction with the Maize
            Mapping Project"

BASE COUNT      215 a      303 c      246 g      186 t
ORIGIN

/Note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT      361 a      490 c      482 g      334 t
ORIGIN

Query Match      5.6%; Score 39.6; DB 11; Length 1667;
Best Local Similarity 47.2%; Pred. No. 1.3;
Matches 120; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 405 CGTCGATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGTT 464
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 979 CGCGCGCGCGCAGCTGACTACGTGCGCGGACCATCAGCGCGCTGGAGAGTCTGTGAA 1038

QY 465 GAACCCGAATCGTATTGAGACGCGTTTGTATTATCGCGCGCTCGCGCGCAATACGG 524
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1039 GGCGCGCGCGCGCAGCTCCCGTGTTCGTGAGCGGCGGCTCCGCGGCGCACCGCGT 1098

QY 525 CGACACGAGCGGTGATTTTCGCCCTCGCGGTTTCGCGGCAAGCTGATCTGTTC 584
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1099 CTTCAAGCGCTCGCGCTCGCGCGCGCGGCTATTTCGTGCGGAGCGCGTGTCTC 1158

QY 585 GCTGTGTGGTTTCGCGCGCAGCAGCATTTGTCAGCGCGCTCCAGCCCAAGGTGTCG 644
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1159 CTGCGCGCGCGCGCGCGCGCGCTGTCACAGCTCTCGCGCATCTCGCGGACGAGTT 1218

QY 645 CTGGATCAACGCTG 658
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1219 CGAGCTCACCATGG 1232

RESULT 9
AY104298/c
LOCUS      AY104298      950 bp      mRNA      linear      HTC 25-MAY-2002
DEFINITION Zea mays PC0095355 mRNA sequence.
ACCESSION  AY104298
VERSION     AY104298.1 GI:21207376
KEYWORDS   HTC.
SOURCE      Zea mays.
ORGANISM   Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 950)
            Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitesitt, M.S.,
            Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
            Maize Mapping Project/DuPont Consensus Sequences for Design of
            Overgo Probes
            Unpublished (2002)
            2 (bases 1 to 950)
            Coe, E.C.
            Direct Submission
            Submitted (25-APR-2002) Maize Mapping Project, University of
            Missouri, Columbia, MO 65211, USA
            Location/Qualifiers
            1. .950
            /organism="Zea mays"
            /db_xref="MaizeDB:635429"
            /db_xref="taxon:4577"
            /clone="PC0095355"
            /clone_lib="Maize Mapping Project/DuPont Consensus
            Library"
            /note="this sequence is part of a project of EST
            assemblies resulting from the application of public
            contigs to seed DuPont contigs; this resource was
            assembled by DuPont as part of a collaboration for the
            overgo addressing of BACs in conjunction with the Maize
            Mapping Project"

BASE COUNT      215 a      303 c      246 g      186 t
ORIGIN
```

```

Query Match      5.5%; Score 39; DB 11; Length 950;
Best Local Similarity 46.5%; Pred. No. 1.6;
Matches 166; Conservative 0; Mismatches 185; Indels 6; Gaps 1;

QY 349 TTGGCGGTTTCGGCGTGGCCACTGACACCGCAACCGGTGGGGTGGAGGTGAGCGTC 408
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 392 TTGGCCCTGGAGCGCGGAGAGTCCGATGGCGCCCTGGAGAGCTTGCCTGGTTATGGTC 333
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 409 GATAAGCAGCGGGTTCGGGTAAGCCCATGTTGATGCAATCGTCTGACCTGCTGTAAC 468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 332 CTCTCGTAGTAGCAGAGTGGAGCCGAGTGGACGTAGCAGGTGCTGGCGAGTGGATC 273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 469 CGGATCGGTATTGGACGGGTTTGTGTTTATCGCGCGGTGCGCGCGCAATACGCGCAC 528
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 272 TCGAAGCGCGCGTGGCTCGTTCGAAGCTGTAGGAGTGACGAGTCCGGGATGAGCCCC 213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 529 ACCGACGGTGGATTTTCGCGCTGGCGGTTCGCGCAAGCCTGATCTGTTCCCGCTG 588
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 212 TTGGGAGCGGCTACTTGGGACAGAGTGTCTCCCGCGCGCTGAGGAGCGC-----G 159
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 589 GTGGGTTTCGCGCAGCAGCATTTCTACGCCCGCTGTCACGCCCAAGGTTGCGCGCTG 648
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 158 GCGACGCGTGGCGCGCGCGCGCGGAGGAGGACGAGGACACGAGCGCGCGCGG 99
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 649 ATCAAGCTGCTGGCGAGTGTGATGACCGCATTTGCCCATCAACTGATGTTGATG 705
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 GGCAGCGGAGACATGGCGGTTTGGCACTTTCGAGTTTGGGGGAAAAAGAGTTCTGGTG 42
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AW787854/c
LOCUS      AW787854
DEFINITION 945004G09.X1 945 - Mixed adult tissues from Walbot lab, same as 707
            (SK) Zea mays cDNA, mRNA sequence.
ACCESSION  AW787854
VERSION     AW787854.1 GI:7844632
KEYWORDS    EST.
SOURCE      Zea mays.
ORGANISM    Zea mays.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
            1 (bases 1 to 557)
            Walbot, V.
            Maize ESTs from various cDNA libraries sequenced at Stanford
            University
            Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 945004 row: G column: 09.
FEATURES
    source
        1. 557
            /organism="Zea mays"
            /cultivar="W23"
            /db_xref="taxon:4577"
            /clone_lib="945 - Mixed adult tissues from Walbot lab,
            same as 707 (SK)"
            /tissue_type="tassel, kernal, silk, husk, root, leaf"
            /dev_stage="fully-grown"
            /lab_host="DH10B"
            /note="Organ: tassel, kernal, silk, husk, root, leaf;
            vector: pBAD10; Site_1: EcoRI; cDNA library from fully
            differentiated maize tissues from an active Mutator plant.
            Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,
            root, leaf). Unidirectionally cloned. New library number
            given to library 707 for additional sequencing."

BASE COUNT      81 a 215 c 161 g 100 t

```

```

ORIGIN
Query Match      5.5%; Score 38.8; DB 10; Length 557;
Best Local Similarity 48.2%; Pred. No. 1.6;
Matches 109; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 349 TTGGCGGTTTCGGCGTGGCCACTGACACCGCAACCGGTGGGGTGGAGGTGAGCGTC 408
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 331 TTGGCCCTGGAGCGCGGAGAGTCCGAGATGGCGCCCTGGAGAGCTTGCCTGGTTATGGTC 272
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 409 GATAAGCAGCGGGTTCGGGTAAGCCCATGTTGATGCAATCGTCTGACCTGCTGTAAC 468
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 271 CTCTCGTAGTAGCAGAGTGGAGCCGAGTGGACGTAGCAGGTGCTGGCGAGTGGATC 212
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 469 CGGATCGGTATTGGACCGGTTTGTGTTTATCGCGCGGTGCGCGCGCAATACGCGCAC 528
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 TCGAAGCGCGCGTGGCTCGTTCGAAGCTGTAGGAGTGACCGAGTCCGGGATGAGCCCC 152
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 529 ACCGACGGTGGATTTTCGCGCTGGCGGTTCGCGCAAGCCTGA 574
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 TTGGGAGCGGCTACTTGGGACAGAGTGTTCGCGCGCGCTGA 106
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
BG975548/c
LOCUS      BG975548
DEFINITION 602845228F1 NCI_CGAP Mam4 Mus musculus cDNA clone IMAGE:4980895 5',
            mRNA sequence.
ACCESSION  BG975548
VERSION     BG975548.1 GI:14363185
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
            1 (bases 1 to 903)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10981 row: g column: 08
            High quality sequence stop: 250.
FEATURES
    Location/Qualifiers
        1. .903
            /organism="Mus musculus"
            /strain="NMRI"
            /db_xref="taxon:10090"
            /clone="IMAGE:4980895"
            /clone_lib="NCI CGAP Mam4"
            /tissue_type="tumor, gross tissue"
            /dev_stage="5 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
            Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
            Library constructed by Life Technologies. Investigators
            providing samples: Lothar Hennighausen/Priscilla Furth,
            NIH Reference for transgenic model: Li et al., Cell Growth
            and Differentiation 7, 3-11 (1996)."

BASE COUNT      373 a 167 c 297 g 66 t
ORIGIN
Query Match      5.3%; Score 38; DB 13; Length 903;
Best Local Similarity 55.2%; Pred. No. 3.2;
Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

```

Qy 127 CTGCTGCTTTAATTTCTGACGCTCTTTTGTTCATCGCGGCACCTTGGCGGTGATCTT 186
Db 560 CTGCTGCTCTAGTCGCGTCTTGTTCGCGCGCGTCTTGTCTGTCGCGCTC 501
Qy 187 TTGTCATGCGCGCGGATCGTCTCATATATATGCTGCGGTGGCATCGCTTACCTG 246
Db 500 TGCTGCTCGCTGCGCGGCTCGTCTCTCTCTTTTCCGCGGTTCGCCGCTGCGTC 441
Qy 247 TTATGCTTGGCGCT 260
Db 440 TTTGCTTTTGTCT 427

RESULT 12
LOCUS CNS01HIJ 804 bp DNA linear GSS 01-JUN-2001
DEFINITION Anopheles gambiae GSS SP6 end of clone 08N07 of Notredame library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL144476
VERSION AL144476.1 GI:7002638
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 804)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source
1. .804
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="08N07"
/clone_lib="Notredame1"
/notes="end : SP6"
BASE COUNT 181 a 257 c 207 g 148 t 11 others
ORIGIN
Query Match 5.3%; Score 37.8; DB 17; Length 804;
Best Local Similarity 48.1%; Pred. NO. 3.6;
Matches 137; Conservative 1; Mismatches 143; Indels 4; Gaps 1;

Qy 412 AAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGATCTGCTGACCTGTTGAACCCG 471
Db 660 ATGAAGCAGGCGCGCTGGAAGATTATCGTCACCATGCTGCGGTGACCTGGCTTAACCCG 601
Qy 472 AATGCGTATTGGACGGTTTGTGTTTATCGCGGGTCTCGCGCGCAATACGGGACCC 531
Db 600 CATGTGTATCTGACACCTTCTGCTGCTGGGAGCTGCGCGGTCACTGGAGGTTGAG 541
Qy 532 GGAAGGTGATTTTCGCGCTGCGCGCTTCGCGCGCAAGCTGATCTGTTTCCGCTGGTG 591
Db 540 CCGAAACGCTGGTTTGGCTCGGTACGGTCAAGCGCTCTTCTCTGTTGTTCTTCGCTC 481
Qy 592 GGTTCGCGGCAGCAGCATTTGTCACGCCCGCTGTCAGCCCGCCAGGTGGCGCTGGATC 651
Db 480 CATTCGCGCGCATG----TGGCGCGCGCGCTGCTGTCACGCGCAAGCCGCGCATCAT 425

```

```

Qy 652 AACGTGCTGTCAGTTGTGATGCGCATTCGCCATCAAACTG 696
Db 424 AATACCTGTCGCGCTGGTATGTTATCGCTTCAGCTG 380

RESULT 13
LOCUS BJ295166/c 370 bp mRNA linear EST 09-APR-2002
DEFINITION BJ295166 Y. Ogiwara unpublished cDNA library, Wh_SL Triticum
aestivum cDNA clone whs12a03 3', mRNA sequence.
ACCESSION BJ295166
VERSION BJ295166.1 GI:20109847
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 370)
AUTHORS Ogiwara,Y. and Murai,K.
TITLE Expressed genes in Triticum aestivum
JOURNAL Unpublished (2002)
COMMENT Contact: Tadaasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source
1. 370
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whs12a03"
/clone_lib="Y. Ogiwara unpublished cDNA library, Wh_SL"
/tissue_type="seed DPA30"
/dev_stage="Peekes' scale 11.3"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown under
hydroponic conditions at UC Davis, salt stressed for 12
hours, and for 7 days, then dissected and frozen (Akhunov
in J Dvorak Lab). Total RNA was prepared from sheath
tissue, equal quantities of RNA were pooled from the two
samples, polyA was purified from the pooled RNA, a cDNA
library was made, and the cDNA clones were in vivo
excised to give pBluescript phagemids in the TJ Clouse lab
at the University of California, Riverside (Akhunov, Chin
, Choi, Clouse, Fenton, Kianian, Otto, Simons, Zhang).
Plasmid DNA preparations and DNA sequencing were
performed in the OD Anderson lab (all other authors)."
BASE COUNT 63 a 127 c 85 g 93 t
ORIGIN
Query Match 5.3%; Score 37.6; DB 13; Length 370;
Best Local Similarity 50.0%; Pred. NO. 3.1;
Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 264 GGCAGCGAAAGACGCATGACAAACAGGTGAAGCGCCACAGATCATTTGAAGAACAGA 323
Db 368 GCGCGCCAGGACGACCGTTGGAGAGACGCGCTCGGCCAAGGACGCCCGATGGAGAC 309
Qy 324 ACCAACCGTGCCTGATGACAGCTTTTGGCGGTTCGGCGGTGCGCACTGACACCGCAA 383
Db 308 GCGCGAGGCGGCAAGGACACGCGCTGGGAGACGCGCGGCGGCAAGGATAAGGCCAA 249
Qy 384 CCGGTGCGGTGGAGGTGAGGTGATGAGCGGTTCGAGCGGTTCGAGTAAAGCCCATGTTAT 443
Db 248 CAGGGGTACGAGAGGTGAAGGAGAGGTTTGGGAGGTGAAGGACAAAGTCAACCGGCG 189
Qy 444 GGCATATCG 451

```



Db 188 GGCAGCG 181

RESULT 14  
LOCUS BH843100/c  
DEFINITION TC3-54P23.TV TC3 Trypanosoma cruzi genomic clone TC3-54P23, DNA sequence.  
ACCESSION BH843100  
VERSION BH843100.1 GI:21411012  
KEYWORDS GSS.  
SOURCE Trypanosoma cruzi.  
ORGANISM Trypanosoma cruzi.  
REFERENCE Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma; Schizotrypanum.  
1 (bases 1 to 451)  
AUTHORS Myler, P.J., Agarwal, G., Fazelinia, G., Mack, J., Marty, A., Munden, H., Nelson, S., Pantony, M., Rinta, J., Robertson, L., Seyler, A., Sisk, E., Stuart, K., Vogt, C., Worthey, E., El-Sayed, N.M., Ghedin, E. and Andersson, B.  
TITLE Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing  
JOURNAL Unpublished (2001)  
COMMENT Contact: Peter Myler  
Seattle Biomedical Research Institute  
4 Nickerson Street, Seattle, WA 98109, USA  
Tel: 206 284 8846  
Fax: 206 284 0313  
Email: mylerp@sbri.org  
Clones are derived from the Trypanosoma cruzi CL-Brener BAC library TC3. For clone availability, please contact Dr. Bjorn Andersson at Uppsala University (bjorn.andersson@genpat.uu.se).  
Seq primer: T7  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..451

/organism="Trypanosoma cruzi"

/strain="CL Brener"

/db\_xref="taxon:5693"

/clone="TC3-54P23"

/clone\_lib="TC3"

/notes="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed for Uppsala University by Marie-Christine Le Paslier in the laboratory of Denis Le Paslier at the Centre d'Etude du Polymorphisme Humain (CEPH), Paris, France. Briefly, Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained from Dr. Franco da Silveira) was partially digested with Hin dIII. High molecular weight fragments were ligated in pBelOBAC11 digested with Hin dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome."

BASE COUNT 89 a 154 c 99 g 109 t

## ORIGIN

Query Match 5.2%; Score 37.2; DB 17; Length 451;  
Best Local Similarity 50.0%; Pred. No. 4.4;  
Matches 93; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 490 TTTGTGTTATTCGGCGCGTTCGGCGCAATACCGCGACACCGGACGTGGATTTCGCC 549

Db 399 TTTGGGCGACGGCGCTCTGACCGCGAGTTGGCGGTGGGATCTTCTCCG 340

QY 550 GCTGGCGGTTCCGGCGACGCTGATCTGGTTCCGCTGGGTTCGGCGGACGACGA 609

Db 339 ACGGCGGTTGATGGGACCTCAAAGCGGCATCTCGGCGCATGTGTTTCGACGCTCAACT 280

QY 610 TTGTCACCCCGCTGTCCAGCCCAAGGTGGCGCTGATCAACGTCGTCTGTCGAGTT 669

Db 279 CGCAGAGAGTGGCGGAATCGGCCCGCAGTGGACCATGACGACGCTTGAGGACATC 220

QY 670 GTGATG 675

Db 219 CTGCTG 214

RESULT 15  
LOCUS CNS06D94/c

DEFINITION T7 end of clone AROAA011D12 of library AROAA from strain CBS 732 of Zygosaccharomyces rouxii, genomic survey sequence.

ACCESSION AL393518

VERSION AL393518.1 GI:12143776

KEYWORDS GSS.

SOURCE Zygosaccharomyces rouxii.

ORGANISM Zygosaccharomyces rouxii.

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Zygosaccharomycetales; Saccharomycetaceae; Zygosaccharomycetes.

1 (bases 1 to 644)

AUTHORS Souciet, J.L., Aigle, M., Artiguenave, P., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrien, P., Lepingle, A., Liorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 644)

AUTHORS de Montigny, J., Straub, M., Potier, S., Tekala, F., Dujon, B., Wincker, P., Artiguenave, P. and Souciet, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 8.

JOURNAL Zygosaccharomyces rouxii

MEDLINE FEBS Lett. 487 (1), 52-55 (2000)

PUBMED 20584718

REFERENCE 3 (bases 1 to 644)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (06-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

source

Location/Qualifiers

1..644

/organism="Zygosaccharomyces rouxii"

/strain="CBS 732"

/db\_xref="taxon:4956"

/clone="AROAA011D12"

/clone\_lib="AROAA"

/note="end : T7"

BASE COUNT 128 a 235 c 88 g 104 t 89 others

## ORIGIN

Query Match 5.2%; Score 37.2; DB 17; Length 644;

Best Local Similarity 32.7%; Pred. No. 5;

Matches 92; Conservative 50; Mismatches 139; Indels 0; Gaps 0;

QY 351 GGGCGGTTTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 410

Db 588 GGGGCGCTCKGSGGKKGKGTGTGGGGGCGGGGGGGGGGGGGGGGGGGGGGG 529

QY 411 TAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTGAACCC 470

Db 528 KCGCGGGCGGCGGKKGKKGKGGGGGKGGKGTGKGGGGGGGGGGGGGGGGGGGGGG 469

QY 471 GAATCGTATTTCGACGCGGTTTCTGTTTATTCGCGCGCGGCGGCGCAATACGCGGACAC 530

```

Db 468 TGGKKKGGGKBBKKKKKGTNTGCTTTTGGTGBBKKKKKGTGKTGGTGTGGGKTGG 409
Qy 531 CGGACGGTGGATTTCGCCGCTGGCGCGTTCGGCGCAAGCCTGATCTGGTTCCCGCTGGT 590
Db 408 KGCTCGSGKTGKGGGTGCGTGGGGTTSKGBDKKKKTCGCGGGTGGGGTSGK 349
Qy 591 GGGTTTCGGCGCAGCAGCAATTGTACGCCCGCTGTCCAGCC 631
Db 348 GKTGKGGTCAAGBBYBKTTGKYBKKGKGGGGGNGCC 308

```

Search completed: April 27, 2003, 00:51:19  
 Job time : 955.401 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 17:06:23 ; Search time 156.463 Seconds  
(without alignments)  
10233.535 Million cell updates/sec

Title: US-09-105-117k-1\_COPY\_1016\_1726

Perfect score: 711

Sequence: 1 atgggtgatcatgaaatctt...-.....aactgatgttgagggttag 711

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq 101002:\*

1: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID82/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID82/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	822	22 AAF71779	Corynebacterium gl
2	711	100.0	822	23 AAS96098	C. glutamicum gene
3	711	100.0	2374	18 AAT96816	DNA encoding LysG.
4	711	100.0	349980	22 AAG68528	C glutamicum codin
5	708	99.6	708	22 AAG68420	C glutamicum codin
6	317.6	44.7	1568	22 AAH45375	C. thermoaminogene
7	60.4	8.5	636	21 AAAS2691	Escherichia coli y
8	57	8.0	15239	17 AAT33536	BCG deletion regio
9	57	8.0	4403765	22 AAI99683	Mycobacterium tube

10	57	8.0	4411529	22 AAI99682	Mycobacterium tube
11	47.4	6.7	7521	23 AAS71378	DNA encoding novel
12	47.4	6.7	7521	23 AAS94251	DNA encoding novel
13	39	5.5	993	22 AAF71777	Corynebacterium gl
14	39	5.5	993	23 AAS96096	C. glutamicum gene
15	37.6	5.3	4403765	22 AAI99683	Mycobacterium tube
16	37.6	5.3	4411529	22 AAI99682	Mycobacterium tube
17	37	5.2	1155	22 AAF26425	Pseudomonas sp hea
18	36.8	5.2	1920	23 ABL29295	Drosophila melanog
19	36.8	5.2	6029	23 ABL17830	Drosophila melanog
20	36.8	5.2	6454	23 ABL29284	Drosophila melanog
21	36.8	5.2	14395	23 ABL29294	Drosophila melanog
22	36.6	5.1	2842	21 AAD00334	Rice raffinose syn
23	36	5.1	1068	22 AAG65404	C glutamicum codin
24	36	5.1	349980	22 AAH68525	C glutamicum codin
25	35	4.9	5828	21 AAA14845	DNA encoding tolue
26	34.8	4.9	7158	24 ABK74880	Bacillus lichenifo
27	34.6	4.9	423	23 AAS93033	DNA encoding novel
28	34.6	4.9	6611	17 AAT09311	Mycobacteriophage
29	34.6	4.9	6611	18 AAT70491	Mycobacteriophage
30	34.6	4.9	6611	18 AAT66129	Mycobacteriophage
31	34.6	4.9	6611	18 AAT51223	Nhei-D fragment of
32	34.4	4.8	459	22 AAH36182	Human colon cancer
33	34.4	4.8	3318	15 AAQ68194	Tn5099-10 transpon
34	34.4	4.8	4372	18 AAT91360	Orf virus genomic
35	34.2	4.8	125401	22 AAD17186	Streptomyces nous
36	34	4.8	300	20 AAZ12866	Human gene express
37	34	4.8	515	20 AAZ17675	Human gene express
38	34	4.8	2048	21 AAF16325	Human prostate can
39	34	4.8	5224	20 AAZ32022	Human METH1 relate
40	34	4.8	524	22 AAC90079	L05390 cDNA clone.
41	34	4.8	5488	22 AAH18214	Human cDNA sequenc
42	33.8	4.8	2005	24 ABL95574	Human angiotensinis
43	33.8	4.8	2005	24 ABL88085	Human PRO7223 cDNA
44	33.8	4.8	2024	21 AAA07697	Human collectin en
45	33.8	4.8	2181	22 AAI60628	Human polynucleoti

## ALIGNMENTS

RESULT 1

AAAF71779

ID AAF71779 standard; DNA; 822 BP.

XX AAF71779;

AC AAF71779;

XX 30-APR-2001 (first entry)

DT Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:53.

DE Corynebacterium glutamicum; metabolic pathway protein; MP protein;

XX fine chemical production; microorganism; organic acid; nucleoside;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX Corynebacterium glutamicum.

OS Corynebacterium glutamicum.

XX WO200100843-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00923.

XX 25-JUN-1999; 99US-0141031.

XX 01-JUL-1999; 99DE-1030476.

XX 02-JUL-1999; 99US-0142101.

XX 08-JUL-1999; 99DE-1031415.

XX 08-JUL-1999; 99DE-1031418.

XX 08-JUL-1999; 99DE-1031419.

XX 08-JUL-1999; 99DE-1031420.

XX 08-JUL-1999; 99DE-1031424.



PD 13-SEP-2001.  
 PF 22-DEC-2000; 2000WO-1B02035.  
 PR 09-MAR-2000; 2000US-187970P.  
 PR 23-JUN-2000; 2000US-0606740.  
 XX (BADI ) BASF AG.  
 PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;  
 PI Lee H, Hwang B;  
 XX WPI; 2001-582269/65.  
 DR P-PSDB; AAU71888.  
 XX  
 PT Nucleic acids encoding metabolic pathway proteins from *Corynebacterium*  
 PT glutamicum, useful for producing methionine and lysine in  
 PT *Corynebacterium* and *Brevibacterium* -  
 XX  
 PS Disclosure; Page 214-215; 316pp; English.  
 XX  
 CC The present invention relates to the isolation of novel *Corynebacterium*  
 CC glutamicum genes encoding metabolic pathway (MP) proteins  
 CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention  
 CC include enzymes involved in the lysine and methionine biosynthetic  
 CC pathways. The polynucleotide sequences of the invention can be used  
 CC for the large-scale production and/or modulation of expression of  
 CC fine chemicals such as lysine and methionine. The sequences of the  
 CC invention may be used to identify *C. glutamicum* and related organisms  
 CC e.g. *C. diphtheriae* in a subject to detect diphtheria.  
 CC AA596073-AAS96132 represent *C. glutamicum* genes encoding the novel  
 CC metabolic pathway proteins of the invention.  
 XX  
 SQ Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;

Query Match 100.0%; Score 711; DB 23; Length 822;  
 Best Local Similarity 100.0%; Pred. No. 2.8e-206;  
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCAATACAGTCTGCTTTGGGGCCAGCTTTTACTGTCC 60  
 DB 92 ATGGTGATCATGGAATCTTCAATACAGTCTGCTTTGGGGCCAGTCTTTTACTGTCC 151  
 QY 61 ATCGGACGCGAGATGTACTGTGTATTAACAAGGAATTAAGCGCAAGGACTCAATCGG 120  
 DB 152 ATCGGACGCGAGATGTACTGTGTATTAACAAGGAATTAAGCGCAAGGACTCAATCGG 211  
 QY 121 GTTCTTCTGCTGTGTTAATTTCTGACGCTTTTGTTCATCGCCGGCACTTGGGCGTT 180  
 DB 212 GTTCTTCTGCTGTGTTAATTTCTGACGCTTTTGTTCATCGCCGGCACTTGGGCGTT 271  
 QY 181 GATCTTTTGTCCAAATGCCGCGCGATCGTCTCGATATTATGCGTGGGTGGCATCGCT 240  
 DB 272 GATCTTTTGTCCAAATGCCGCGCGATCGTCTCGATATTATGCGTGGGTGGCATCGCT 331  
 QY 241 TACCTGTTATGTTTGGCTCATGCGAGCAAGCGCATGACAAACAAGTGAAGCG 300  
 DB 332 TACCTGTTATGTTTGGCTCATGCGAGCAAGCGCATGACAAACAAGTGAAGCG 391  
 QY 301 CCACAGATCATTTGAAGAAACAGAAACAAACCGTGCCTGATGACACCCCTTTGGGCGGTTTCG 360  
 DB 392 CCACAGATCATTTGAAGAAACAGAAACAAACCGTGCCTGATGACACCCCTTTGGGCGGTTTCG 451  
 QY 361 GCGGTGGCCACTGACACCGGCAACCGGTGCGGTGGAGGTGAGGTGATGATGACAGCGG 420  
 DB 452 GCGGTGGCCACTGACACCGGCAACCGGTGCGGTGGAGGTGAGGTGATGATGACAGCGG 511  
 QY 421 GTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGAATGCGTAT 480  
 DB 512 GTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGAATGCGTAT 571  
 QY 481 TTGAGCGGTTTGTGTTTATCGGCGGCTGCGCGCGCAATACGGCGACCGGACGGTGG 540

DB 572 TTGACGCGTTTGTGTTTATCGCGCGGTTCGGCGCAATACGGCGACACCGGACGGTGG 631  
 QY 541 ATTTTCGCGCTGGCGGTTCGGCGCAAGCTGATCTGTTCCGCTGGGTTCGGC 600  
 DB 632 ATTTTCGCGCTGGCGGTTCGGCGCAAGCTGATCTGTTCCGCTGGGTTCGGC 691  
 QY 601 GCAGCAGCATTTGTCAACGCGCTGTCCAGCGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 660  
 DB 692 GCAGCAGCATTTGTCAACGCGCTGTCCAGCGCCCAAGGTGTGGCGCTGGATCAACGTCGTC 751  
 QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTGATGGGTTAG 711  
 DB 752 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTGATGGGTTAG 802

## RESULT 3

AAT96816

ID AAT96816 standard; DNA; 2374 BP.

XX AC AAT96816;

XX 12-MAR-1998 (first entry)

XX DNA encoding LysG, LysE and ORF3 from *Corynebacterium glutamicum*.

DE LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;  
 KW Microbial production; amino acid; animal feed additive; ds.  
 XX

XX *Corynebacterium glutamicum*.

FH Key Location/Qualifiers  
 FT CDS complement (82..954)

FT /\*tag= a  
 FT /label= LysG

FT 1016..1726  
 FT /\*tag= b

FT /label= LysE  
 FT CDS complement (1723..2373)

FT /\*tag= c  
 FT /label= orf3

XX DE19548222-A1.

XX 26-JUN-1997.

XX 22-DEC-1995; 95DE-1048222.

XX 22-DEC-1995; 95DE-1048222.

XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.

XX Eggeling L, Sahn H, Vrljic M;

XX WPI; 1997-333867/31.

XX P-PSDB; AAW37714-16.

XX Increasing microbial production of amino acids, especially lysine -  
 PT by improving export carrier activity or corresponding gene  
 PT expression, also new export and regulatory genes from  
 PT *Corynebacterium*

XX Claim 23 and 26; Page -; 16pp; German.

XX This DNA, isolated from *Corynebacterium glutamicum*, contains the LysG,

CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory  
 CC protein and an export protein, respectively. Microbial production of  
 CC amino acids (A) is improved by increasing the export-carrier activity  
 CC and/or the export gene expression in a microorganism that produces (A).  
 CC The method is specifically used to increase production of lysine, as  
 CC used as an animal feed additive. Other (A) are variously useful as  
 CC pharmaceuticals, condiments and intermediates for fine chemicals.  
 CC This method increases the amount of (A) secreted into the culture medium.  
 CC Export of (A) has been found to depend on a single gene.

CC NB. This sequence has been created from the information given in table 2  
 CC of the specification.

XX  
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;  
 Query Match 100.0%; Score 711; DB 18; Length 2374;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-206;  
 Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGGCGAGTCTTTTACTGTGCC 60  
 DB 1016 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGGCGAGTCTTTTACTGTGCC 1075  
 QY 61 ATCGGACCGCAGAAATGTAATGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 120  
 DB 1076 ATCGGACCGCAGAAATGTAATGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 1135  
 QY 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGGGGTT 180  
 DB 1136 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGGGGTT 1195  
 QY 181 GATCTTTTGTCCAAATGCGCGCGGATCGTGTCTCGATATTATGCGTGGGGTGGCATCGCT 240  
 DB 1196 GATCTTTTGTCCAAATGCGCGCGGATCGTGTCTCGATATTATGCGTGGGGTGGCATCGCT 1255  
 QY 241 TACCTGTTTATGTTTGGCTCATGCGGAGGAGCGCATGACAAACAAGGTGGAAGCG 300  
 DB 1256 TACCTGTTTATGTTTGGCTCATGCGGAGGAGCGCATGACAAACAAGGTGGAAGCG 1315  
 QY 301 CCACAGATCATTAAGAAACAGAACCAACCGTGCCTGATGACACGCCCTTTGGGGGTTTCG 360  
 DB 1316 CCACAGATCATTAAGAAACAGAACCAACCGTGCCTGATGACACGCCCTTTGGGGGTTTCG 1375  
 QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 420  
 DB 1376 GCGGTGGCCACTGACACGCGCAACCGGTGGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 1435  
 QY 421 GTTTCGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAAACCCGAATGCGTAT 480  
 DB 1436 GTTTCGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAAACCCGAATGCGTAT 1495  
 QY 481 TTGGACGCGTTGTTTATCGCGCGGCTGCGCGCAATACGCGCACCGGAGCGTGG 540  
 DB 1496 TTGGACGCGTTGTTTATCGCGCGGCTGCGCGCAATACGCGCACCGGAGCGTGG 1555  
 QY 541 ATTTTCGCGCTGGCGGCTTGGCGCAAGCTGATCTGTTCCCGCTGGTGGGTTTCGGC 600  
 DB 1556 ATTTTCGCGCTGGCGGCTTGGCGCAAGCTGATCTGTTCCCGCTGGTGGGTTTCGGC 1615  
 QY 601 GCAGCAGCATTTGACGCGCGCTGTCAGCCCGCCAGAGGTGCGCTGATCAACGTCGTC 660  
 DB 1616 GCAGCAGCATTTGACGCGCGCTGTCAGCCCGCCAGAGGTGCGCTGATCAACGTCGTC 1675  
 QY 661 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 711  
 DB 1676 GTGGCAGTTGTGATGACCGCATTTGCCATCAAACTGATGTTGATGGGTTAG 1726

## RESULT 4

AAH68528/c

ID AAH68528 standard; DNA; 349980 BP.

XX

AC

XX

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 7063.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

OS

XX

PN

XX BP1108790-A2.

PD

XX 20-JUN-2001.

XX

XX 18-DEC-2000; 2000BP-0127688.

XX

XX 16-DEC-1999; 99JP-0377484.

PR

PR 07-APR-2000; 2000JP-0159162.

PR

PR 03-AUG-2000; 2000JP-0280988.

XX

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX

XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;

PI

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX

XX WPI; 2001-376931/40.

XX

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT

PT mutation point of a gene, measuring expression of a gene, analysing

PT

PT expression profile or pattern of a gene and identifying homologous gene

PT

PS Disclosure; SEQ ID NO: 7063; 246pp + Sequence Listing; English.

XX

XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a nucleic acid described  
 CC in the exemplification of the invention.

CC

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

XX

SQ Sequence 349980 BP; 80727 A; 91049 C; 97618 G; 80586 T; 0 other;

Query Match

Best Local Similarity 100.0%; Score 711; DB 22; Length 349980;

Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGGCGAGTCTTTTACTGTGCC 60

DB 128953 ATGGTGATCATGGAATCTTCATTACAGTCTGCTTTTGGGGGCGAGTCTTTTACTGTGCC 128894

QY

QY 61 ATCGGACCGCAGAAATGTAATGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 120

DB

DB 128893 ATCGGACCGCAGAAATGTAATGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 128834

QY

QY 121 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGGCGTT 180

DB

DB 128833 GTTCTTCTCGTGTGTTTAAATTTCTGACGCTCTTTTGTTCATCGCGGACCTTTGGGCGTT 128774

QY

QY 181 GATCTTTTGTCCAAATGCGCGCGGATCGTGTCTCGATATTATGCGTGGGGTGGCATCGCT 240

DB

DB 128773 GATCTTTTGTCCAAATGCGCGCGGATCGTGTCTCGATATTATGCGTGGGGTGGCATCGCT 128714

QY

QY 241 TACCTGTTTATGTTTGGCTCATGCGGAGGAGCGCATGACAAACAAGGTGGAAGCG 300

DB

DB 128713 TACCTGTTTATGTTTGGCTCATGCGGAGGAGCGCATGACAAACAAGGTGGAAGCG 128654

QY

QY 301 CCACAGATCATTAAGAAACAGAACCAACCGTGCCTGATGACACGCCCTTTGGGCGGTTTCG 360

DB

DB 128653 CCACAGATCATTAAGAAACAGAACCAACCGTGCCTGATGACACGCCCTTTGGGCGGTTTCG 128594

QY

QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 420

DB

DB 128593 GCGGTGGCCACTGACACGCGCAACCGGTGGGGTGGAGGTGAGCGTGCATTAAGCAGCGG 128534

QY

QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTGTGACCTGTTGAACCCGAATGCGTAT 480



DB 128533 GTTGGGTAAGCCCATGTTGATGGCAATCGTCTGACCTGTTGAACCGAATCGGTAT 128474  
QY 481 TTGGACGCGTTTGTGTTTATCGCGCGGTCGGCGGCATACGGCGACACCGGAGCGTGG 540  
DB 128473 TTGGACGCGTTTGTGTTTATCGCGCGGTCGGCGGCATACGGCGACACCGGAGCGTGG 128414  
QY 541 ATTTTGGCGCGTGGCGGTTTCGGCAAGCTGATCTGGTTCCCGCTGGTGGGTTTCGGC 600  
DB 128413 ATTTTGGCGCGTGGCGGTTTCGGCAAGCTGATCTGGTTCCCGCTGGTGGGTTTCGGC 128354  
QY 601 GCAGCAGCAATGTCAGCGCGGTCGACGCCCAAGGTTGGCGCTGGATCAACGTCGTC 660  
DB 128353 GCAGCAGCAATGTCAGCGCGGTCGACGCCCAAGGTTGGCGCTGGATCAACGTCGTC 128294  
QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGGGTTAG 711  
DB 128293 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGGGTTAG 128243

RESULT 5  
AAH68420  
ID AAH68420 standard; DNA; 708 BP.

XX AC AAH68420;  
XX DT 26-SEP-2001 (first entry)  
XX C glutamicum coding sequence fragment SEQ ID NO: 3455.  
XX DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX KW organic acid synthesis; ds.  
XX OS Corynebacterium glutamicum.

XX FN EP1108790-A2.  
XX PD 20-JUN-2001.  
XX PF 18-DEC-2000; 2000EP-0127688.  
XX PR 16-DEC-1999; 99JP-0377484.  
XX PR 07-APR-2000; 2000JP-0159162.  
XX PR 03-AUG-2000; 2000JP-0280988.  
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX DR WPI; 2001-376931/40.  
XX DR P-PSDB; AAG93201.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
XX FT mutation point of a gene, measuring expression of a gene, analysing  
XX FT expression profile or pattern of a gene and identifying homologous gene

XX PS Claim 1; SEQ ID NO: 3455; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein  
XX CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
XX CC are useful for identifying the mutation point of a gene derived from a  
XX CC mutant of coryneform bacterium, measuring expression amount and  
XX CC analysing the expression profile or expression pattern of a gene derived  
XX CC from Coryneform bacterium, and identifying a homologue of a gene derived  
XX CC from coryneform bacterium. Coryneform bacteria are useful for producing  
XX CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
XX CC particularly L-lysine. The present sequence is a nucleic acid described  
XX CC in the exemplification of the invention.  
XX CC Note: The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from the  
XX CC European Patent Office.

XX SQ Sequence 708 BP; 134 A; 173 C; 221 G; 180 T; 0 other;  
Query Match 99.8%; Score 708; DB 22; Length 708;  
Best local similarity 100.0%; Pred. No. 2.1e-205;  
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCTGATCATGGAATCTTTTATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60  
DB 1 ATGCTGATCATGGAATCTTTTATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60  
QY 61 ATCGGACCGCAGAAATGCTGCTGATTAACAAGGAATTAAGCGCGAGGACTCATTTGCG 120  
DB 61 ATCGGACCGCAGAAATGCTGCTGATTAACAAGGAATTAAGCGCGAGGACTCATTTGCG 120  
QY 121 GTTCTTCTCGTGTGTTTAAATTTCTGAGCTCTTTTGTTCATCCCGGACCTTTGGGCGTT 180  
DB 121 GTTCTTCTCGTGTGTTTAAATTTCTGAGCTCTTTTGTTCATCCCGGACCTTTGGGCGTT 180  
QY 181 GATCTTTTGTCCAATGCGCGCGATCGTCTGATATTATGCTGGGTGGCATCGCT 240  
DB 181 GATCTTTTGTCCAATGCGCGCGATCGTCTGATATTATGCTGGGTGGCATCGCT 240  
QY 241 TACCTGTTATGTTGCTGCTCATGGCAGGAAAGAGCCCATGACAAAGGTGGAAGCG 300  
DB 241 TACCTGTTATGTTGCTGCTCATGGCAGGAAAGAGCCCATGACAAAGGTGGAAGCG 300  
QY 301 CCACAGATCAATTGAAGAAACAGAACCAACGTCGCCGATGACACGCTTTGGCGGTTTCG 360  
DB 301 CCACAGATCAATTGAAGAAACAGAACCAACGTCGCCGATGACACGCTTTGGCGGTTTCG 360  
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGCGAGTGAGCTGCGATAGCAGCGG 420  
DB 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGCGAGTGAGCTGCGATAGCAGCGG 420  
QY 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAACCCGAATGCGTAT 480  
DB 421 GTTTGGGTAAAGCCCATGTTGATGCAATCGTCTGACCTGTTGAACCCGAATGCGTAT 480  
QY 481 TTGGACGCGTTTGTGTTTATCGCGCGGTCGGCGCGCAATACCGCGACACCGGAGTGG 540  
DB 481 TTGGACGCGTTTGTGTTTATCGCGCGGTCGGCGCGCAATACCGCGACACCGGAGTGG 540  
QY 541 ATTTTGGCGCGTGGCGGTTTCGGCAAGCTGATCTGTTCCCGCTGGGTTTCGGC 600  
DB 541 ATTTTGGCGCGTGGCGGTTTCGGCAAGCTGATCTGTTCCCGCTGGGTTTCGGC 600  
QY 601 GCAGCAGCAATGTCAGCGCGGTCGTCAGCCCCCAAGGTTGGCGCTGGATCAACGTCGTC 660  
DB 601 GCAGCAGCAATGTCAGCGCGGTCGTCAGCCCCCAAGGTTGGCGCTGGATCAACGTCGTC 660  
QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGGGTTAG 708  
DB 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGGGTTAG 708

RESULT 6  
AAH45375  
ID AAH45375 standard; DNA; 1568 BP.  
XX AC AAH45375;  
XX DT 11-SEP-2001 (first entry)  
XX C. thermoamniogenes lysin biosynthetic enzyme lyse DNA.  
XX DE Heat-resistant; lysin biosynthesis; enzyme; coryneform;  
XX KW aspartate-semialdehyde dehydrogenase; lyseE; ds.  
XX OS Corynebacterium thermoamniogenes.  
XX PN JP2001120270-A.  
XX



QY 199 GCGCGCATGCTGCATATATATGCGTGGGGTGGCATCGCTTACCTGTTATGTTTGGC 258  
 Db |||||  
 193 TCGCGGTGTTGCTGGCGCTGGTCACCTGGCGCGGTAGCCCTTCTGCTGTGATGTT 252  
 QY 259 GTCATGCGCAGCAAGACGCGCATGACAAACAAGTGA 296  
 Db |||||  
 253 TTTGGCGCTTTTAAACAGCAATGAGCAGTAATATTGA 290

RESULT 8  
 AAT33536/c  
 ID AAT33536 standard; DNA; 15239 BP.  
 XX  
 AC AAT33536;  
 XX  
 DT 15-FEB-1998 (first entry)  
 DE BCG deletion region 2 and flanking sequences.

XX BCG delta 2; virulence; avirulence; attenuation; gene deletion;  
 KW mycobacteria; vaccine; infection; marker; ss.  
 XX

OS Mycobacterium bovis strain BCG.

FH Key Location/Qualifiers  
 FT misc\_feature 3382..14071  
 FT /\*tag= a  
 FT /note= "BCG delta 1 deletion region"

PN WO9625519-Al.

XX 22-AUG-1996.

XX 15-FEB-1996; 96WO-US01938.

XX 17-FEB-1995; 95US-0390878.

PA (PATH-) PATHOGENESIS CORP.

PI Mahairas GG, Stover CK;

DR WPI; 1996-393419/39.

XX Detecting markers for avirulence in Mycobacterium - used in  
 PT production of vaccines against bacterial infection, and to detect  
 PT bacterial infection

PS Example 1; Fig 2; 66pp; English.

XX This DNA sequence comprises Mycobacterium bovis BCG deletion  
 CC sequence BCGdelta2. A specific genetic deletion of this region  
 CC results in an avirulence phenotype of the mycobacterium. 2 Other  
 CC deletion regions (see AAT33535 and AAT33537) have also been detected.  
 CC Identification involved screening a BCG cosmid library with a  
 CC radiolabeled probe obtained following DNA subtraction between  
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.  
 CC The deletions provide useful markers for the identification of an  
 CC avirulent, or a virulent, mycobacterial phenotype. Determination  
 CC of avirulence requires the detection of the presence or absence of  
 CC the deletion; the deletions are detected either by detecting the  
 CC presence or absence of deletion junctions (see AAT33538-46), or by  
 CC detecting the presence or absence of the sequences contained within  
 CC the deletion. Deletion polypeptides are used as components of  
 CC immunological assays and in vaccines.

XX Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;

Query Match 8.0%; Score 57; DB 17; Length 15239;  
 Best Local Similarity 53.1%; Pred. No. 2.1e-06;  
 Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTGATGGCAATGCTGCTGACCTGGTTGAACCGCAATGCTATTTGGACGCT 490  
 |||||

Db 4792 AGGTCTCTGTGACCTGTGCGGCATTCACGTTCTCAACCCACACGCTTACCTCGACACCG 4733  
 QY 491 TTGTCGTTTATCGGCGCGCTCGGCGCGCAATACGGCGACACCGGCGGTGGATTTTCGCGG 550  
 Db |||||  
 4732 TCGTGTGCTAGCGCGCTGGCCAAACGAGCACAGCA---COAGCGCTGGCTGTTTCGGCC 4676  
 QY 551 CTGGCGCGTTCGCGCAAGCCTGATCTCGTTCCCGCTCGTGGGTTCGCGCGCAGCAGCAT 610  
 Db |||||  
 4675 TCGGCGCGGTACAGCCAGTGGGTATGTTGCCACCTCGGGTTCGAGCGCGCGGT 4616  
 QY 611 TGTCACGCGCGCTGTCCAGCCCAAGGTGTGGCGCTGGATCAAGCTGCTGTCGTCAGTTG 670  
 Db |||||  
 4615 TCGCGCGGCTGTTTACCAACCCCGCTCGTGGAGAACTCTCGACGCGCTGATCGCGTCA 4556  
 QY 671 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 703  
 Db |||||  
 4555 TGATGTTGCGCTGGGAATCTCGCTGACCGTGA 4523

RESULT 9

AAI99683

ID AAI99683 standard; DNA; 4403765 BP.

XX

AC AAI99683;

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
 XX

OS Mycobacterium tuberculosis.

PN US6294328-B1.

XX 25-SEP-2001.

PF 24-JUN-1998; 98US-0103840.

PR 24-JUN-1998; 98US-0103840.

PA (GENO-) INST GENOMIC RES.

PI Fleischmann RD, White OR, Fraser CM, Venter JC;

DR WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ

PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 8.0%; Score 57; DB 22; Length 4403765;

Best Local Similarity 53.1%; Pred. No. 2.9e-05;

Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTATGGCAATCGTCTGACCTGGTTGAACCGGAATGCGTATTTGGAGCGGT 490  
Best Local Similarity 53.1%; Pred. No. 2.9e-05;  
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTATGGCAATCGTCTGACCTGGTTGAACCGGAATGCGTATTTGGAGCGGT 490  
Db 2220334 AGTCTCTGGTACCTGTGCGGCATTCACGTTCTCAACCCACACGCTTACCTCGACACCG 2220393

QY 491 TTGCTGTTTATCGCGCGCTCGCGCGCAATACGGCGACACCGAGCGTGGATTTTCGCCG 550  
Db 2220394 TCGTGTCTAGGCGCGCTGCGCAACGAGCAGCGA--CCAGCGTGGCTGTTTCGCC 2230450

QY 551 CTGCGCGCTTCGCGCGCAAGCCTGATCTGTTCCCGCTGGTTCGCGCGCAGCAGCAT 610  
Db 2230451 TCGCGCGCTTCACAGCCAGTGGTATGTTTCGCGCGCTTCGCGCGCGGT 2230510

QY 611 TGTACGCGCGCTTCGCGCGCAAGTTCGCGCGCTGGATCAAGTCTGTCGCGAGTTG 670  
Db 2230511 TCGCGCGCTTCACAGCCAGTGGTATGTTTCGCGCGCTTCGCGCGCGGT 2230570

QY 671 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 703  
Db 2230571 TGATGTTGCGTGGGATCTCGCTGACCGTGA 2230603

RESULT 11  
ID AAS71378 standard; DNA; 4411529 BP.  
XX AAS71378;  
AC AAS71378;  
XX  
DT 15-JAN-2002 (first entry)  
XX  
DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
XX  
KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
KW variation; epidemiology; patient treatment; epidemic monitoring; ds.  
XX  
OS Mycobacterium tuberculosis.  
XX  
PN US6294328-B1.  
XX  
PD 25-SEP-2001.  
XX  
PF 24-JUN-1998; 98US-0103840.  
XX  
PR 24-JUN-1998; 98US-0103840.  
XX  
PA (GENO-) INST GENOMIC RES.  
XX  
XX Fleischmann RD, White OR, Fraser CM, Venter JC;  
XX WPI; 2001-647261/74.  
XX  
XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ -  
XX  
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
XX  
CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen.  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at [seqdata.uspto.gov/sequence.html?DocID=6294328B1](http://seqdata.uspto.gov/sequence.html?DocID=6294328B1).

SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
Query Match 8.0%; Score 57; DB 22; Length 4411529;  
Best Local Similarity 53.1%; Pred. No. 2.9e-05;  
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTATGGCAATCGTCTGACCTGGTTGAACCGGAATGCGTATTTGGAGCGGT 490  
Db 2220334 AGTCTCTGGTACCTGTGCGGCATTCACGTTCTCAACCCACACGCTTACCTCGACACCG 2220393

QY 491 TTGCTGTTTATCGCGCGCTCGCGCGCAATACGGCGACACCGAGCGTGGATTTTCGCCG 550  
Db 2220394 TCGTGTCTAGGCGCGCTGCGCAACGAGCAGCGA--CCAGCGTGGCTGTTTCGCC 2230450

QY 551 CTGCGCGCTTCGCGCGCAAGCCTGATCTGTTCCCGCTGGTTCGCGCGCAGCAGCAT 610  
Db 2230451 TCGCGCGCTTCACAGCCAGTGGTATGTTTCGCGCGCTTCGCGCGCGGT 2230510

QY 611 TGTACGCGCGCTTCGCGCGCAAGTTCGCGCGCTGGATCAAGTCTGTCGCGAGTTG 670  
Db 2230511 TCGCGCGCTTCACAGCCAGTGGTATGTTTCGCGCGCTTCGCGCGCGGT 2230570

QY 671 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 703  
Db 2230571 TGATGTTGCGTGGGATCTCGCTGACCGTGA 2230603

RESULT 11  
ID AAS71378 standard; cDNA; 7521 BP.  
XX AAS71378;  
AC AAS71378;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #7182.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG07191.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 7182; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 7521 BP; 1896 A; 1856 C; 2100 G; 1669 T; 0 other;

Query Match 6.7%; Score 47.4; DB 23; Length 7521;  
Best Local Similarity 48.4%; Pred. No. 0.0012;  
Matches 132; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 29 GTCTGCTTTTGGGGCCAGTCTTTTACTGTCTCCATCGGACCGCAGAAATGCTACTGGTGATTA 88

DB 4494 GTCTTGCACTTGGGGCGGCTATGCTTACCGCTCGGTCCACAAATGCTTTTGATGA 4553

QY 89 AACAGGAATTAAGCGGAGGACTCATTCGCGTTCTTCTCGTGTGTTTAATTTCTGACG 148

DB 4554 ATCAGGGCATACGTCGTGAGTACCAATATGATGTCCTTACTTTTGTCTATCAGCGATT 4613

QY 149 TCTTTTGTTCATCGCCGACCTTGGCGGTGATCTTTTGTCCATGCCGCGGATCG 208

DB 4614 TGGTCTCTGATTGGCGCGGATTTTGGTGACGCGCTTATGATGCGCGGTGGT 4673

QY 209 TGCTCGATATTATGCGCTGGGCTGTCATCGCTTACCTGTATGTTTGGCTCATGGCAG 268

DB 4674 TGCTGGCGCTGTCACCTGGGCGCGGTAACTTCTTCTGTGCAACAGACTACCGGG 4733

QY 269 CGAAGAGCGCATGACAAACAGGTGGAGCGC 301

DB 4734 GGAAGAGAGCGGAAGAGGAGGAGGAGAGC 4766

## RESULT 12

AAS94251  
ID AAS94251 standard; cDNA; 7521 BP.

XX AC AAS94251;

XX DT 13-FEB-2002 (first entry)

XX DE DNA encoding novel human diagnostic protein #30055.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX XX WPI; 2001-639362/73.

XX DR P-PSDB; ABG30064.

XX FT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX PS Claim 1; SEQ ID No 30055; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 7521 BP; 1896 A; 1856 C; 2100 G; 1669 T; 0 other;

Query Match 6.7%; Score 47.4; DB 23; Length 7521;  
Best Local Similarity 48.4%; Pred. No. 0.0012;

Matches 132; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 29 GTCTGCTTTTGGGGCCAGTCTTTTACTGTCCATCGGACCGCAGAAATGCTACTGGTGATTA 88

DB 4494 GTCTTGCACTTGGGGCGGCTATGCTTACCGCTCGGTCCACAAATGCTTTTGATGA 4553

QY 89 AACAGGAATTAAGCGGAGGACTCATTCGCGGTCTTCTCGTGTGTTTAATTTCTGACG 148

DB 4554 ATCAGGGCATACGTCGTGAGTACCAATATGATGTCCTTACTTTTGTCTATCAGCGATT 4613

QY 149 TCTTTTGTTCATCGCCGACCTTGGCGGTGATCTTTTGTCCATGCCGCGGATCG 208

DB 4614 TGGTCTCTGATTGGCGCGGATTTTGGTGACGCGCTTATGATGCGCGGTGGT 4673

QY 209 TGCTCGATATTATGCGCTGGGCTGTCATCGCTTACCTGTATGTTTGGCTCATGGCAG 268

DB 4674 TGCTGGCGCTGTCACCTGGGCGCGGTAACTTCTTCTGTGCAACAGACTACCGGG 4733

QY 269 CGAAGAGCGCATGACAAACAGGTGGAGCGC 301

DB 4734 GGAAGAGAGCGGAAGAGGAGGAGGAGAGC 4766

## RESULT 13

AAP71777/c

ID AAP71777 standard; DNA; 993 BP.

XX AC AAP71777;

XX DT 30-APR-2001 (first entry)

XX DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.

XX KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;

XX KW fine chemical production; microorganism; organic acid; nucleoside;

XX KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

XX KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

XX KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX OS Corynebacterium glutamicum.

PN WO200100843-A2.  
XX 04-JAN-2001.  
XX 23-JUN-2000; 2000WO-IB00923.  
XX 25-JUN-1999; 99US-0141031.  
XX 01-JUL-1999; 99DE-1030476.  
XX 02-JUL-1999; 99US-0142101.  
XX 08-JUL-1999; 99DE-1031415.  
XX 08-JUL-1999; 99DE-1031418.  
XX 08-JUL-1999; 99DE-1031419.  
XX 08-JUL-1999; 99DE-1031420.  
XX 08-JUL-1999; 99DE-1031424.  
XX 08-JUL-1999; 99DE-1031428.  
XX 08-JUL-1999; 99DE-1031434.  
XX 08-JUL-1999; 99DE-1031435.  
XX 08-JUL-1999; 99DE-1031443.  
XX 08-JUL-1999; 99DE-1031453.  
XX 08-JUL-1999; 99DE-1031457.  
XX 08-JUL-1999; 99DE-1031465.  
XX 08-JUL-1999; 99DE-1031478.  
XX 08-JUL-1999; 99DE-1031510.  
XX 08-JUL-1999; 99DE-1031541.  
XX 08-JUL-1999; 99DE-1031573.  
XX 08-JUL-1999; 99DE-1031592.  
XX 08-JUL-1999; 99DE-1031632.  
XX 08-JUL-1999; 99DE-1031634.  
XX 08-JUL-1999; 99DE-1031636.  
XX 08-JUL-1999; 99DE-1032125.  
XX 08-JUL-1999; 99DE-1032126.  
XX 08-JUL-1999; 99DE-1032130.  
XX 08-JUL-1999; 99DE-1032186.  
XX 08-JUL-1999; 99DE-1032206.  
XX 08-JUL-1999; 99DE-1032227.  
XX 08-JUL-1999; 99DE-1032228.  
XX 08-JUL-1999; 99DE-1032229.  
XX 08-JUL-1999; 99DE-1032230.  
XX 08-JUL-1999; 99DE-1032922.  
XX 14-JUL-1999; 99DE-1032926.  
XX 14-JUL-1999; 99DE-1032928.  
XX 14-JUL-1999; 99DE-1033004.  
XX 14-JUL-1999; 99DE-1033005.  
XX 14-JUL-1999; 99DE-1033006.  
XX 12-AUG-1999; 99US-0148613.  
XX 27-AUG-1999; 99DE-1040764.  
XX 27-AUG-1999; 99DE-1040765.  
XX 27-AUG-1999; 99DE-1040766.  
XX 27-AUG-1999; 99DE-1040832.  
XX 31-AUG-1999; 99DE-1041378.  
XX 31-AUG-1999; 99DE-1041379.  
XX 31-AUG-1999; 99DE-1041380.  
XX 31-AUG-1999; 99DE-1041394.  
XX 31-AUG-1999; 99DE-1041396.  
XX 03-SEP-1999; 99DE-1042076.  
XX 03-SEP-1999; 99DE-1042077.  
XX 03-SEP-1999; 99DE-1042079.  
XX 03-SEP-1999; 99DE-1042086.  
XX 03-SEP-1999; 99DE-1042087.  
XX 03-SEP-1999; 99DE-1042088.  
XX 03-SEP-1999; 99DE-1042095.  
XX 03-SEP-1999; 99DE-1042124.  
XX 03-SEP-1999; 99DE-1042129.  
XX 09-NAR-2000; 2000US-0187970.  
XX (BADI ) BASF AG.  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX WPI; 2001-137957/14.  
XX P-PSDB; AAB79658.  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX Claim 3; Page 226-228; 1737pp; English.  
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC MP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polyketides and enzymes.  
XX Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;  
SQ Query Match 5.5%; Score 39; DB 22; Length 993;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTG 39  
DB 39 ATGGTGATCATGGAATCTTCATTACAGGTCTGCTTTTG 1  
RESULT 14  
AAS96096/c  
ID AAS96096 standard; DNA; 993 BP.  
XX AC AAS96096;  
XX DT 26-FEB-2002 (first entry)  
XX DE C. glutamicum gene #21 encoding metabolic pathway protein.  
XX KW Metabolic pathway protein; MP; lysine biosynthesis pathway;  
XX KM methionine biosynthesis pathway; large-scale production of fine chemical;  
XX KW Corynebacterium diphtheriae; diphtheria; ds.  
XX OS Corynebacterium glutamicum.  
XX WO200166573-A2.  
XX 13-SEP-2001.  
XX 22-DEC-2000; 2000WO-IB02035.  
XX 09-MAR-2000; 2000US-187970P.  
XX 23-JUN-2000; 2000US-0606740.  
XX (BADI ) BASF AG.  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;  
XX Lee H, Hwang B;  
XX WPI; 2001-582269/65.  
XX P-PSDB; AAB71886.  
XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium  
PT glutamicum, useful for producing methionine and lysine in  
PT Corynebacterium and Brevibacterium -  
XX Disclosure; Page 207-208; 316pp; English.  
XX The present invention relates to the isolation of novel Corynebacterium  
CC glutamicum genes encoding metabolic pathway (MP) proteins  
CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention  
CC include enzymes involved in the lysine and methionine biosynthetic  
CC pathways. The polynucleotide sequences of the invention can be used  
CC for the large-scale production and/or modulation of expression of  
CC fine chemicals such as lysine and methionine. The sequences of the  
CC invention may be used to identify C. glutamicum and related organisms  
CC e.g. C. diphtheriae in a subject to detect diphtheria.

CC AAS96073-AAS96132 represent C. glutamicum genes encoding the novel  
CC metabolic pathway proteins of the invention.

XX  
SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

Query Match 5.5%; Score 39; DB 23; Length 993;

Best Local Similarity 100.0%; Pred. No. 0.18;

Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAAATCTTCATTACAGGCTGCTTTTG 39

DB 39 ATGGTGATCATGGAAATCTTCATTACAGGCTGCTTTTG 1

## RESULT 15

AAI99683/c

ID AAI99683 standard; DNA; 4403765 BP.

XX AC AAI99683;

XX DT 15-JAN-2002 (first entry)

XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;

XX KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX OS Mycobacterium tuberculosis.

XX PN US6294328-B1.

XX PD 25-SEP-2001.

XX PF 24-JUN-1998; 98US-0103840.

XX PR 24-JUN-1998; 98US-0103840.

XX PA (GENO-) INST GENOMIC RES.

XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises  
PT determining the nucleotide sequence of the strain at positions in the  
PT genome corresponding to positions where M. tuberculosis strains CDC  
PT 1551 and H37Rv differ

PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX CC The invention relates to evaluating strain variation within and between  
CC different populations of the tuberculosis bacterial pathogen.  
CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
CC nucleotide sequence of the first strain at positions in the complete  
CC sequence of the genome that correspond to positions that differ in the  
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
CC M. tuberculosis and has valuable application in the fields of  
CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 5.3%; Score 37.6; DB 22; Length 4403765;

Best Local Similarity 47.5%; Pred. No. 23;

Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 430 AAGCCCATGTTGATCGATCGCTGACCTGGTTGACCCGAATCGATTGTCGCG 489

DB 720443 AAGGACCTGCGCTGCGCTCGAGCGGCGGCGGAGCTGCCGCGCCCGCTCGGT 720384

QY 490 TTTGTGTTTATCGGCGGCGTTCGGCGCAATACGGCGCACCGGACGGTGGATTTTCGCC 549

DB 720383 GGGCTGCTGTTACCTGCAACGGGCGCGGACGACGATGTTTCGGGGTCAACGACACGAC 720324

QY 550 GCTGGCGCGTTCGGCGCAAGCCCTGATCTGTTCCCGCTGGTGGGTTTCGGCGCGCAGCAGCA 609

DB 720323 GCGTCGACGATCGAGGACCTGCTGGGCGGGATTCGCTGCGCGGTTTCTTCGCCCGCGG 720264

QY 610 TTGTACGCGCGCTGTTCAGCCGCCCAAGGTGTGGCGCTGGATCAACGTCGTCGTGGC 665

DB 720263 GAGATCGCGCGCGTTCGGCGGCCACACGCGTTGACGCGTTTACCGCGTCGATGGC 720208

Search completed: April 26, 2003, 22:40:54

Job time : 7840.46 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 00:51:48 ; Search time 80.2974 Seconds  
(without alignments)  
9634.930 Million cell updates/sec

Title: US-09-105-117K-1\_COPY\_1016\_1726

Perfect score: 711

Sequence: 1 atggtgatcatggaatctt.....aactgatgttgagggttag 711

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	711	100.0	822	9	US-09-746-660A-51
C 2	711	100.0	3309400	9	US-09-738-626-1
3	708	99.6	708	9	US-09-738-626-1455
4	57	8.0	593	9	US-09-894-844-11
C 5	39	5.5	993	9	US-09-746-660A-47
C 6	37.2	5.2	424	10	US-09-960-352-12392
7	36	5.1	1058	9	US-09-738-626-439
8	36	5.1	3309400	9	US-09-738-626-1
9	35.4	5.0	671	9	US-10-184-644-346
10	35.4	5.0	671	9	US-10-184-634-346
11	35	4.9	5828	9	US-09-430-029-1
12	34.8	4.9	7158	10	US-09-974-300-2171
C 13	34.4	4.8	4372	9	US-09-796-679-4
C 14	34	4.8	2048	10	US-09-925-300-760
C 15	33.8	4.8	2930	10	US-09-745-763-198
C 16	33.8	4.8	88421	9	US-09-976-059-1
17	33.6	4.7	1075	10	US-09-864-761-19241
18	33.6	4.7	1403	10	US-09-864-761-2513
C 19	33.6	4.7	1635	10	US-09-864-761-20241

C 20	33.6	4.7	1973	10	US-09-864-761-3471
21	33.4	4.7	412	9	US-09-918-995-36145
C 22	33.2	4.7	160	10	US-09-923-876-582
23	33	4.6	446	10	US-09-864-761-20699
24	32.8	4.6	443	9	US-09-918-995-13829
C 25	32.8	4.6	443	9	US-09-918-995-20282
26	32.8	4.6	479	9	US-09-918-995-31047
C 27	32.8	4.6	562	9	US-09-918-995-37138
28	32.8	4.6	2209	10	US-09-726-397A-2
C 29	32.8	4.6	2366	10	US-09-726-397A-3
30	32.8	4.6	2568	9	US-09-738-626-15
31	32.8	4.6	3955	9	US-10-219-248-4
32	32.8	4.6	3955	9	US-10-219-247-4
33	32.8	4.6	3955	10	US-09-855-732-4
34	32.8	4.6	4702	10	US-09-880-107-1545
35	32.8	4.6	62944	10	US-09-954-456-2257
36	32.6	4.6	520	9	US-10-184-644-332
37	32.6	4.6	520	9	US-10-184-634-332
C 38	32.6	4.6	1164	9	US-09-712-363-110
39	32.6	4.6	3627	9	US-09-927-827-4
40	32.6	4.6	7356	9	US-09-927-827-1
C 41	32.4	4.6	497	9	US-09-918-995-27219
42	32.4	4.6	591	10	US-09-833-381-112
43	32.4	4.6	3333	9	US-09-991-053-5
44	32.4	4.6	3333	9	US-09-957-187-5
45	32.4	4.6	3498	9	US-09-991-053-3

ALIGNMENTS

RESULT 1  
US-09-746-660A-51  
; Sequence 51, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CP2  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 51  
; LENGTH: 822  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)..(799)  
; OTHER INFORMATION: RXA01394

US-09-746-660A-51

Query Match 100.0%; Score 711; DB 9; Length 822;  
Best Local Similarity 100.0%; Pred. No. 7.9e-232;  
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60  
DB 92 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 151  
QY 61 ATCCGACCGCAGATGACTGCTGATTAACAAGCAATTAAGCCGAGGACTCAATTGCG 120  
DB 152 ATCCGACCGCAGATGACTGCTGATTAACAAGCAATTAAGCCGAGGACTCAATTGCG 211  
QY 121 GTTCTTCTCGTGTTAAATTTCTGAGCTCTTTTGTTCATCGCGCACCTTGGGCGTT 180  
DB 212 GTTCTTCTCGTGTTAAATTTCTGAGCTCTTTTGTTCATCGCGCACCTTGGGCGTT 271  
QY 181 GATCTTTTGTCCATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 240  
DB 272 GATCTTTTGTCCATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 331  
QY 241 TACCTGTTATGTTGCGCTCATGCGCAGCAAGACGCGCATGACAAACAAGGTGGAAGCG 300  
DB 332 TACCTGTTATGTTGCGCTCATGCGCAGCAAGACGCGCATGACAAACAAGGTGGAAGCG 391  
QY 301 CCACAGATCATTTGAAGAAACAGAACCGTGCCTGATGACACGCTTTGGGCGTTGCG 360  
DB 392 CCACAGATCATTTGAAGAAACAGAACCGTGCCTGATGACACGCTTTGGGCGTTGCG 451  
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGAGGTGAGCGTGCATAGCAGCGG 420  
DB 452 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGAGGTGAGCGTGCATAGCAGCGG 511  
QY 421 GTTTGGGTAAAGCCCATGTCATGCGCAATCGTCTGACCTGTTGAACCCGATGCGTAT 480  
DB 512 GTTTGGGTAAAGCCCATGTCATGCGCAATCGTCTGACCTGTTGAACCCGATGCGTAT 571  
QY 481 TTGACGCGTTTGTGTTTATCGGCGCGTTCGCGCGCAATAACGCGACACCGGACGTTG 540  
DB 572 TTGACGCGTTTGTGTTTATCGGCGCGTTCGCGCGCAATAACGCGACACCGGACGTTG 631  
QY 541 ATTTTTCGCGCTGGCGGTTCGCGCGAAGCTGATCTGTTCCGCTGGTGGTTTCGCG 600  
DB 632 ATTTTTCGCGCTGGCGGTTCGCGCGAAGCTGATCTGTTCCGCTGGTGGTTTCGCG 691  
QY 601 GCAGCAGATTTGTCACGCGCGCTGTCAGCGCCCAAGGTGTGCGCTGATCAACGTCGTC 660  
DB 692 GCAGCAGATTTGTCACGCGCGCTGTCAGCGCCCAAGGTGTGCGCTGATCAACGTCGTC 751  
QY 661 GTGGCAGTTGTGATGACCGCATTTGCCATCAACTGATGTTGATGGGTTAG 711  
DB 752 GTGGCAGTTGTGATGACCGCATTTGCCATCAACTGATGTTGATGGGTTAG 802

## RESULT 2

US-09-738-626-1/c  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 100.0%; Score 711; DB 9; Length 3309400;  
Best Local Similarity 100.0%; Pred. No. 6.4e-230;  
Matches 711; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 60  
DB 1328953 ATGGTGATCATGGAATCTTCAATACAGGTCTGCTTTTGGGGCCAGTCTTTTACTGTCC 1328894  
QY 61 ATCCGACCGCAGATGACTGCTGATTAACAAGCAATTAAGCCGAGGACTCAATTGCG 120  
DB 1328893 ATCCGACCGCAGATGACTGCTGATTAACAAGCAATTAAGCCGAGGACTCAATTGCG 1328834  
QY 121 GTTCTTCTCGTGTTAAATTTCTGAGCTCTTTTGTTCATCGCGCACCTTGGGCGTT 180  
DB 1328833 GTTCTTCTCGTGTTAAATTTCTGAGCTCTTTTGTTCATCGCGCACCTTGGGCGTT 1328774  
QY 181 GATCTTTTGTCCATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 240  
DB 1328773 GATCTTTTGTCCATGCGCGCGATCGTCTCGATATTATGCGCTGGGTGGCATCGCT 1328714  
QY 241 TACCTGTTATGTTTGGCTCATGCGCAGCAAGACCGCATGACAAACAAGTGGAGCG 300  
DB 1328713 TACCTGTTATGTTTGGCTCATGCGCAGCAAGACCGCATGACAAACAAGTGGAGCG 1328654  
QY 301 CCACAGATCATTTGAAGAAACAGAACCGTGCCTGCGCATGACACGCTTTGGGCGTTGCG 360  
DB 1328653 CCACAGATCATTTGAAGAAACAGAACCGTGCCTGCGCATGACACGCTTTGGGCGTTGCG 1328594  
QY 361 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGAGGTGAGCGTGCATAGCAGCGG 420  
DB 1328593 GCGGTGGCCACTGACACGCGCAACCGGTGCGGTGAGGTGAGCGTGCATAGCAGCGG 1328534  
QY 421 GTTTGGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGTTGAACCCGATGCGTAT 480  
DB 1328533 GTTTGGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGTTGAACCCGATGCGTAT 1328474  
QY 481 TTGGACGCGTTTGTGTTTATCGGCGCGCTCGGCGCAATACGCGCACACCGGACGTTGCG 540  
DB 1328473 TTGGACGCGTTTGTGTTTATCGGCGCGCTCGGCGCAATACGCGCACACCGGACGTTGCG 1328414  
QY 541 ATTTTTCGCGCTGGCGGTTCGCGGCAAGCTGATCTGTTCCGCTGGTGGTTTCGCG 600  
DB 1328413 ATTTTTCGCGCTGGCGGTTCGCGGCAAGCTGATCTGTTCCGCTGGTGGTTTCGCG 1328354  
QY 601 GCAGCAGCATTTGTCACGCGCTGTCAGCCCAAGGTGTGCGCTGGATCAACGTCGTC 660  
DB 1328353 GCAGCAGCATTTGTCACGCGCTGTCAGCCCAAGGTGTGCGCTGGATCAACGTCGTC 1328294  
QY 661 GTGGCAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATGGGTTAG 711  
DB 1328293 GTGGCAGTTGTGATGACCGCATTTGGCCATCAACTGATGTTGATGGGTTAG 1328243

RESULT 3  
US-09-738-626-3455  
; Sequence 3455, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO  
APPLICANT: SENO, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 3455  
LENGTH: 708  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3455

Query Match 99.6%; Score 708; DB 9; Length 708;  
Best Local Similarity 100.0%; Pred. No. 7.7e-231;  
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGATCATGGAATCTTCATTACAGTCTCTTTGGGGCCAGTCTTTTACTGTCC 60  
DB 1 ATGTGATCATGGAATCTTCATTACAGTCTCTTTGGGGCCAGTCTTTTACTGTCC 60

QY 61 ATCGGACGCGAATGTAAGTCTGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 120  
DB 61 ATCGGACGCGAATGTAAGTCTGATTAACAAGGAATTAAGCGGAGGACTCATTTGCG 120

QY 121 GTTCTTCTCGTGTGTTAAATTTCTGACGCTTTTGTTCATCGCGCGGACCTTGGGGGTT 180  
DB 121 GTTCTTCTCGTGTGTTAAATTTCTGACGCTTTTGTTCATCGCGCGGACCTTGGGGGTT 180

QY 181 GATCTTTTGTCCAAATGCGCGCCGATCGTCTCGATATTATCGCTCGGGTGGCATCGCT 240  
DB 181 GATCTTTTGTCCAAATGCGCGCCGATCGTCTCGATATTATCGCTCGGGTGGCATCGCT 240

QY 241 TACTGTATATGTTGGCTCATCGGCGAAAGACGCCCATGACAAACAAGGTGGAAGCG 300  
DB 241 TACTGTATATGTTGGCTCATCGGCGAAAGACGCCCATGACAAACAAGGTGGAAGCG 300

QY 301 CCACAGATCATTTGAAGAACAGAACCAACCGTGGCCGATGACGCTTTGGGGGTTGCG 360  
DB 301 CCACAGATCATTTGAAGAACAGAACCAACCGTGGCCGATGACGCTTTGGGGGTTGCG 360

QY 361 GCGGTGGCCACTGACGCGCAACCGGTGGGTGGAGGTGAGCGTGCATGACGACGCG 420  
DB 361 GCGGTGGCCACTGACGCGCAACCGGTGGGTGGAGGTGAGCGTGCATGACGACGCG 420

QY 421 GTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTGAACCCGGAATGCGTAT 480  
DB 421 GTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTGAACCCGGAATGCGTAT 480

QY 481 TTGACGCGTTTGTGTTATCGGGCGGTGCGCGCAATACGCGGACACCGGACGCGTGG 540  
DB 481 TTGACGCGTTTGTGTTATCGGGCGGTGCGCGCAATACGCGGACACCGGACGCGTGG 540

QY 541 ATTTTTCGCGCTGCGGCTGCGGCAAGCTGATCTGTTCCGCTGGTGGTTCGGC 600  
DB 541 ATTTTTCGCGCTGCGGCTGCGGCAAGCTGATCTGTTCCGCTGGTGGTTCGGC 600

QY 601 GCACGACATTTGTCACGCCCGCTGTCCAGCCCCCAAGGTGTGGCGCTGGATCAACGTCGTC 660

DB 601 GCAGGACATTTGTCACGCCCGTGTCCAGCCCCAAGGTGTGGCGCTGGATCAACGTCGTC 660

QY 661 GTGGCGATTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGT 708

DB 661 GTGGCGATTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATGGGT 708

## RESULT 4

US-09-894-844-11  
Sequence 11, Application US/09894844  
Patent No. US20020176873A1  
GENERAL INFORMATION:  
APPLICANT: Behr, Marcel  
APPLICANT: Small, Peter  
APPLICANT: Schoonik, Gary  
APPLICANT: Wilson, Michael A.  
TITLE OF INVENTION: Molecular Differences Between Species of  
FILE REFERENCE: STAN102CON the M. Tuberculosis Complex  
CURRENT APPLICATION NUMBER: US/09/894,844  
CURRENT FILING DATE: 2001-06-27  
PRIOR APPLICATION NUMBER: 09/318,191  
PRIOR FILING DATE: 1999-05-25  
PRIOR APPLICATION NUMBER: 60/097,936  
PRIOR FILING DATE: 1998-08-25  
NUMBER OF SEQ ID NOS: 137  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 597  
TYPE: DNA  
ORGANISM: Mycobacteria tuberculosis  
US-09-894-844-11

Query Match 8.0%; Score 57; DB 9; Length 597;  
Best Local Similarity 53.1%; Pred. No. 4.9e-09;  
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 431 AGCCCATGTTGATGGCAATCGTCTGACCTGTTGACCCGATGCGTATTTGGAGCGGT 490  
DB 326 AGGTCTGTGTGACCTGTGCGGCATTTACGTTCTCTCAACCCACACGCTCTACCTCGACACCG 385

QY 491 TTGTGTTTATCGGCGCGTTCGCGGCAATACGCGGCAATACGCGGACACCGGATTTTCGCGG 550  
DB 386 TCGTGTGTTGAGCGCGCTGGCCACAGACACAGCA---CCAGCGCTGGCTGTTTCGCGC 442

QY 551 CTGGCGGTTTCGCGGCAAGCCTGATCTGTTCCCGTGGTGGGTTTCGCGCGAGCAGCAT 610  
DB 443 TCGGCGCGTTCACAGCGAGTGGTATGTTCCGACCCCTCGGGTTTCGAGCGCGCGGT 502

QY 611 TGTACGCGCGTGTTCACGCCCCAAGGTGTGGCGTGGATCAACGTCGTCGTCGCGAGTTG 670  
DB 503 TCGCGCGGTGTTACCAACCCCGCTCGTGAGAAATCTCGACGCGCTGTATCGCGGTCA 562

QY 671 TGATGACCGCATTTGGCCATCAAACTGATTTGCA 703

DB 563 TGATGTTGCGTGGGAATCTGCTGACCGTGA 595

## RESULT 5

US-09-746-660A-47/c  
Sequence 47, Application US/09746660A  
Publication No. US20030049804A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schroder, Hartwig  
APPLICANT: Zelder, Oskar  
APPLICANT: Habernauer, Gregor  
APPLICANT: Kim, Jun-Won  
APPLICANT: Lee, Heung-Schick  
APPLICANT: Hwang, Byung-Joon  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING

;; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
;; FILE REFERENCE: BGI-121CP2  
;; CURRENT APPLICATION NUMBER: US/09/746,660A  
;; CURRENT FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 09/606740  
;; PRIOR FILING DATE: 2000-06-23  
;; PRIOR APPLICATION NUMBER: 09/603124  
;; PRIOR FILING DATE: 2000-06-23  
;; PRIOR APPLICATION NUMBER: 60/141031  
;; PRIOR FILING DATE: 1999-06-25  
;; PRIOR APPLICATION NUMBER: 60/142101  
;; PRIOR FILING DATE: 1999-07-02  
;; PRIOR APPLICATION NUMBER: 60/148613  
;; PRIOR FILING DATE: 1999-08-12  
;; PRIOR APPLICATION NUMBER: 60/187970  
;; PRIOR FILING DATE: 2000-03-09  
;; PRIOR APPLICATION NUMBER: DE 19931420.9  
;; PRIOR FILING DATE: 1999-07-08  
;; NUMBER OF SEQ ID NOS: 125  
;; SOFTWARE: PatentIn Vers. 2.0  
;; SEQ ID NO 47  
;; LENGTH: 993  
;; TYPE: DNA  
;; ORGANISM: Corynebacterium glutamicum  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (101)..(970)  
;; OTHER INFORMATION: RXA01393  
US-09-746-660A-47

Query Match 5.5%; Score 39; DB 9; Length 993;  
Best Local Similarity 100.0%; Pred. No. 0.0087;  
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGGTGATCATGGAATCTTATTACAGGTCTGTTTG 39  
Db 39 ATGGTGATCATGGAATCTTATTACAGGTCTGTTTG 1

## RESULT 6

US-09-960-352-12392/c  
;; Sequence 12392, Application US/09960352  
;; Patent No. US20020137139A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Warren, Wesley C.  
;; APPLICANT: Tao, Nengboing  
;; APPLICANT: Byatt, John C.  
;; APPLICANT: Mathialagan, Nagappan  
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
;; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
;; FILE REFERENCE: 16511.006/37-21(10298)C  
;; CURRENT APPLICATION NUMBER: US/09/960,352  
;; CURRENT FILING DATE: 2001-09-24  
;; NUMBER OF SEQ ID NOS: 15112  
;; SEQ ID NO 12392  
;; LENGTH: 424  
;; TYPE: DNA  
;; ORGANISM: Bos taurus  
;; OTHER INFORMATION: Clone ID: 53-LIB3058-002-Q1-K1-F2  
US-09-960-352-12392

Query Match 5.2%; Score 37.2; DB 10; Length 424;  
Best Local Similarity 47.8%; Pred. No. 0.023;  
Matches 108; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 349 TTGGCGCGTTCGGCGTGGCCACTGACACGCGCAACCGGGTGCAGGTGAGGTC 408  
Db 301 TTGGCGTGGACGCGGAGAGTCCGAGATGGCGCCCTGGAGAGCTTCCCGTTATGTC 242  
Qy 409 GATACGACCGGTTTGGGTAAGCCCATTTGATGGCAATCGTGCACCTGTTGAAC 468  
Db 241 CTCCTGTTAGACGAGGTGGAGCCGGAAGTGGACGTAGCAGGTGCTGGCGAGGTGATC 182

Qy 469 CCGAATGCGTATTGTGACGCGTTTGTGTTTATCGCGCGCGTTCGCGCGCAATACGCGCAC 528  
Db 181 TCGAAGCGCGCGTGGCTCGTCGAAGCTGTAGGAGTCCGAGTCCGCGATGAGCCCC 122  
Qy 529 ACCGAGCGGTGGATTTTCGCGCGTGGCGGTTTCGCGGCAAGCCTGA 574  
Db 121 TTCGGGAGGCGGTACTTTGGGACAGAGTCTGTTTCGCGCGCGCGCTGA 76

## RESULT 7

US-09-738-626-439  
;; Sequence 439, Application US/09738626  
;; Publication No. US20020197605A1  
;; GENERAL INFORMATION:  
;; APPLICANT: NAKAGAWA, SATOSHI  
;; APPLICANT: MIZOGUCHI, HIROSHI  
;; APPLICANT: ANDO, SEIKO  
;; APPLICANT: HAYASHI, MIKIRO  
;; APPLICANT: OCHIAI, KEIKO  
;; APPLICANT: YOKOI, HARUHIKO  
;; APPLICANT: TATEISHI, NAKO  
;; APPLICANT: SENOH, AKIHIRO  
;; APPLICANT: IKEDA, MASATO  
;; APPLICANT: OZAKI, AKIO  
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
;; FILE REFERENCE: 249-125  
;; CURRENT APPLICATION NUMBER: US/09/738,626  
;; CURRENT FILING DATE: 2000-12-18  
;; PRIOR APPLICATION NUMBER: JP 99/377484  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: JP 00/159162  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: JP 00/280988  
;; PRIOR FILING DATE: 2000-08-03  
;; NUMBER OF SEQ ID NOS: 7059  
;; SOFTWARE: PatentIn ver. 3.0  
;; SEQ ID NO 439  
;; LENGTH: 1068  
;; TYPE: DNA  
;; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-439

Query Match 5.1%; Score 36; DB 9; Length 1068;  
Best Local Similarity 53.6%; Pred. No. 0.095;  
Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 551 CTGGCGCGTTTCGCGCAAGCTGATCTGGTTCCCGCTGGTGGTTTCGCGCGAGCAGCAT 610  
Db 542 CCGGCGCGTGCACCTGTTCTTCTATCATCGCGCGGACGAGCTCCCGGAAGAAATTA 601  
Qy 611 TGTACGCGCGCTGTCCAGCCCCCAAGTGTGGCGCTGGATCAACGTGCTGCGGAGTTG 670  
Db 602 TTTTCTGGCAGATGGTTCCCTCAACGGCTCGCAATGGCGGACGTCGTCGTTGTTG 661  
Qy 671 TGATGACCGCATTTGCCATC 690  
Db 662 TGATCGTGAATCGGCCTC 681

## RESULT 8

US-09-738-626-1  
;; Sequence 1, Application US/09738626  
;; Publication No. US20020197605A1  
;; GENERAL INFORMATION:  
;; APPLICANT: NAKAGAWA, SATOSHI  
;; APPLICANT: MIZOGUCHI, HIROSHI  
;; APPLICANT: ANDO, SEIKO  
;; APPLICANT: HAYASHI, MIKIRO  
;; APPLICANT: OCHIAI, KEIKO  
;; APPLICANT: YOKOI, HARUHIKO  
;; APPLICANT: TATEISHI, NAKO  
;; APPLICANT: SENOH, AKIHIRO  
;; APPLICANT: IKEDA, MASATO

; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; PRIOR FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: JP 00/159162  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: JP 00/280988  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: Patent in ver. 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 3309400  
 ; TYPE: DNA  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-1

Query Match 5.1%; Score 36; DB 9; Length 3309400;  
 Best Local Similarity 53.6%; Pred. No. 6.6;  
 Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
 QY 551 CTGGCGGCTTCCGCGCAGCCGATGCTGTTCCCGCTGGTGGTTCGCGCAGCAGCAT 610  
 Db 414184 CGGCGCGCTGACCTCGTTCTTCTATCATCGCGCAGCAGCTCCCGCGAAGAAATTA 414243  
 QY 611 TGTACGCGCGCTGTCACGCCCAAGGTGTGGCGCTGGATCAAGCTGCTGGCGAGTTG 670  
 Db 414244 TTTTCTGCAGATGGTTCCTTCAACGGCTCGAATGGCGCAGCAGCTCAACGTGGTGTG 414303  
 QY 671 TGATGACCGCATTTGCCATC 690  
 Db 414304 TGATGTCACAAATCGGCTC 414323

RESULT 9  
 US-10-184-644-346  
 ; Sequence 346, Application US/10184644  
 ; Publication No. US20030044930A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3430R1C227  
 ; CURRENT APPLICATION NUMBER: US/10/184,644  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 346  
 ; LENGTH: 671  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-184-644-346

Query Match 5.0%; Score 35.4; DB 9; Length 671;  
 Best Local Similarity 6.3%; Pred. No. 0.12;  
 Matches 29; Conservative 168; Mismatches 265; Indels 0; Gaps 0;  
 QY 108 AGGACTCATTGCGGTCTCTCGTGTGTTAATTTCTGACGCTTTTGTTCATCGCGG 167  
 Db 32 AVKPPNKPYPFFGTHETAPLPGKDLFFYDKCKDKYKPKRKGPNGLWEIQNPHAS 91

QY 168 CACCTTGGCGGTGATCTTTTGTCCAATGCGCGCCGATCGTCTCGATATTATGCGCTG 227  
 Db 92 YSAPPPVSSSDSEAPANPADGSDADDDDRGVMATATATAASDRMESDSDSKSSD 151  
 QY 228 GGGTGGCATCGTTACCTGTTATGTTTGGCTCATGGCAGCAAGAACCCATGACAAA 287  
 Db 152 NSGLKRTKTPALKMSVSKARKASDLDAQSVSPSEENSESESEKTSQDQFTPEKAA 211  
 QY 288 CAAGGTGAAGCCCAAGATCATTGAGAAACAGAACCAACCGTCCCGCATCACAGCC 347  
 Db 212 VRAPRRGLGRKKKKAPSASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVX 271  
 QY 348 TTTGGCGGTTCGCGGTGCGCATGACACGCGCAACCGGCTGCGTGAGGTGAGCGT 407  
 Db 272 KPPGRKPAEKPLPKPRKPKPERPPSSSSSDSDSDVDRISWKRRDRARRREARR 331  
 QY 408 CGATAACGACGGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTGA 467  
 Db 332 RREQEELRLREQEKEKERRERADRGAEKSGSSGDELREDDPEPVKGRKGRGR 391  
 QY 468 CCCGAATCGCTATTGACCGCTTTGTTTATCGGCGCGCTCGGCGCAATACGCGA 527  
 Db 392 GPSSSDSEPEABLEKAKSAKQSSSTEPARKPKQEKVRPEEKQAKPVKVERTR 451  
 QY 528 CACGACGCGGTGATTTTCGCGCTGCGCGTTCGCGGCAAG 569  
 Db 452 KRSEGFMDRKEKKEPSVEEKLQKLHSEIKALKVDSPDV 493

RESULT 10  
 US-10-184-634-346  
 ; Sequence 346, Application US/10184634  
 ; Publication No. US2003006864A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baker, Kevin P.  
 ; APPLICANT: Chen, Jian  
 ; APPLICANT: Desnoyers, Luc  
 ; APPLICANT: Goddard, Audrey  
 ; APPLICANT: Godowski, Paul J.  
 ; APPLICANT: Gurney, Austin L.  
 ; APPLICANT: Pan, James  
 ; APPLICANT: Smith, Victoria  
 ; APPLICANT: Watanabe, Colin K.  
 ; APPLICANT: Wood, William I.  
 ; APPLICANT: Zhang, Zemin  
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 ; FILE REFERENCE: P3430R1C217  
 ; CURRENT APPLICATION NUMBER: US/10/184,634  
 ; CURRENT FILING DATE: 2002-06-28  
 ; Prior Application removed - See File Wrapper or Palm  
 ; NUMBER OF SEQ ID NOS: 612  
 ; SEQ ID NO 346  
 ; LENGTH: 671  
 ; TYPE: PRT  
 ; ORGANISM: Homo Sapien  
 US-10-184-634-346

Query Match 5.0%; Score 35.4; DB 9; Length 671;  
 Best Local Similarity 6.3%; Pred. No. 0.12;  
 Matches 29; Conservative 168; Mismatches 265; Indels 0; Gaps 0;  
 QY 108 AGGACTCATTGCGGTCTCTCGTGTGTTAATTTCTGACGCTTTTGTTCATCGCGG 167  
 Db 32 AVKPPNKPYPFFGTHETAPLPGKDLFFYDKCKDKYKPKRKGPNGLWEIQNPHAS 91  
 QY 168 CACCTTGGCGGTGATCTTTGTCCAATGCGCGCGATCGTCTGATATTATGCGCTG 227  
 Db 92 YSAPPPVSSSDSEAPANPADGSDADDDDRGVMATATATAASDRMESDSDSKSSD 151  
 QY 228 GGGTGGCATCGTTACCTGTTATGTTTGGCTCATGGCAGCAAGAACCCATGACAAA 287  
 Db 152 NSGLKRTKTPALKMSVSKARKASDLDAQSVSPSEENSESESEKTSQDQFTPEKAA 211

Qy		288	C A G G T G G A A C G C C A C A G A T C A T T G A A G A A C A G A A C A C A C G T C C C G A T G A C A C G C C	347
Db		212	V R A P R R G L G R K K K A P S A D S D K A D S D G A K P E P V A M A R S A S S S S S S D S D V S V K	271
Qy		348	T T T G G G C G G T T C G C G G T G C C A C T G A C A C G C C A A C C G G G T C G G G T G A G G T G A G C G T	407
Db		272	K P P R G R P A E K P I P K P R K R P K P E R P P S S S S S D S D S D E V D R I S E W K R R D E A R R R E A R R	331
Qy		408	C G A T A A C A C G C G G T T T G G T A A G C C C A T G T T G A T G G C A A T C G T C T G A C T T G T T G A A	467
Db		332	R R E Q E E L R L R L E Q E K E K E R R E R A D R G E A E G S G S S G D E L U D E D P V K K R G R K G R K	391
Qy		468	C C G A A T G C G T A T T T G A C C G T T T G T T T A T C G G C G G G T C G G C G C A T A C G C G A	527
Db		392	G P S S S D S E P A E L E R A K K S A K P Q S S T P E A R K P Q K E K R V R P E E K O A K P V K R T R	451
Qy		528	C A C G G A C G G T G G A T T T T C G C G G T G G C G G T T C G C G G C A A G	569
Db		452	K R S G F M D R K R V E K K K P S V E E K L O H S I K F A L K V D S P D V	493

RESULT 11  
US-09-430-029-1  
; Sequence 1, Application US/09430029  
; Patent No. US20020168738A1  
; GENERAL INFORMATION:  
; APPLICANT: Yano, Tetsuma; No. US20020168738A1oto, tsuyoshi; Imamura, Takeshi; Canon K  
; TITLE OF INVENTION: DNA Fragment Carrying Toluene Monooxygenase Gene;  
; TITLE OF INVENTION: Recombinant Plasmid, Transformed Microorganism,  
; TITLE OF INVENTION: Method for Degrading Chlorinated Aliphatic Hydrocarbon  
; TITLE OF INVENTION: Compounds and Aromatic Compounds, and  
; TITLE OF INVENTION: Method for Environmental Remediation  
; FILE REFERENCE: CPO13982US  
; CURRENT APPLICATION NUMBER: US/09/430,029  
; CURRENT FILING DATE: 1999-10-29  
; EARLIER APPLICATION NUMBER: JP P1998-310801  
; EARLIER FILING DATE: 1998-10-30  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1

	Query Match	4.9%;	Score 35;	DB 9;	Length 5828;
	Best Local Similarity	50.3%;	Pred. No. 0.51;		
	Matches	86;	Conservative	0;	Mismatches 85; Indels 0; Gaps
Qy	322	GAACCAACGTCGCCGATGACACGCCTTTGGGGCGGTTCCGGCGGTGGCCACTGACACGCGC	381		
Db	5499	GAAGCGTACGGGATCGCGACCGGAGATGCTGCCCGAAGCGCGCTCCCGCGCGTCCGGCGCGC	5558		
Qy	382	AACCGCGTGGCGGTGGAGGTGACGCTCGATGAAGCAGCGGTTTGGGTAAAGCCCATGTTG	441		
Db	5559	CAACTGCGGTTCTCTGTCGCGACGCGCCATGAACATGCGCGCTGTTCCGAGAAGCGCGTGT	5618		
Qy	442	ATGSGCATCGTGTGACTCTGGTTTGAACCCGAATCGGTATTTTGGACGCGTTTT	492		
Db	5619	GTGGGACACCGCGTTCGCTTCGTGAACCCCGATCCGTTGGCCCGACGACATT	5669		

```

RESULT 12
US-09-974-300-2171
; Sequence 2171, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2171
; LENGTH: 7158
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-2171

```

	Query Match	4.9%;	Score 34.8;	DB 10;	Length 7158;
	Best Local Similarity	54.8%;	Pred. No. 0.66;		
	Matches 69;	Conservative 0;	Mismatches 57;	Indels 0;	Gaps
Qy	450	CGTCTGACCTGGTTGACCCGGAATCGGTATTGGACGCGTTTGTATTATCGCGCGGT	509		
Db	5650	CGGGGTACTTGGCGACCCCGGATTTGACAGCAGACCGCTTTGTCTGCTGATCCGTTTGGT	5708		
Qy	510	CGGCGCGCAATACGGCGACACCGGACGGGTGGGATTTTCGCCGCTGGCGCGCTTCGCCGCGCAAG	569		
Db	5710	CCGCGGGGAACCGGATGTACCGCACCGGCGCATCTTGCCCGCAGGCGCCAGCAGCGCTCT	5768		
Qy	570	CCTGAT	575		
Db	5770	CTGGAT	5775		

```

RESULT 13
US-02-796-679-4/c
; Sequence 4, Application US/09796679
; Publication No. US20030013076A1
; GENERAL INFORMATION:
; APPLICANT: Robinson, Anthony J
; APPLICANT: Lyttle, David J
; APPLICANT: University of Otago
; TITLE OF INVENTION: Parapoxvirus vectors
; FILE REFERENCE: 23607 MRB
; CURRENT APPLICATION NUMBER: US/09/796,679
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/155,421
; PRIOR FILING DATE: 1998-09-29

```

PRIOR APPLICATION NUMBER: PCT/NZ97/00040  
PRIOR FILING DATE: 1997-03-27  
PRIOR APPLICATION NUMBER: NZ 286284  
PRIOR FILING DATE: 1996-03-29  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 4372  
TYPE: DNA  
ORGANISM: Orf virus strain NZ-2  
US-09-796-679-4

Query Match 4.8%; Score 34.4; DB 9; Length 4372;  
Best Local Similarity 43.8%; Pred. No. 0.7;  
Matches 149; Conservative 0; Mismatches 191; Indels 0; Gaps 0;  
QY 364 GTGGCCACTGCACACGCAACCGGTGGGTGGAGTGCAGGTGATTAAGCAGGGGTT 423  
DB 1493 GAGCGCGGGCTGGGGCGAGGGGTGCGCTTCGAGTCCGCCCTCGGAGCACCAGATC 1434  
QY 424 TGGGTAAAGCCCATGTTGATGCAATCTGCTGACCTGGTTGAACCCCAATGCGTATTG 483  
DB 1433 CGGCTCGGGTGGATGCGGTGACGTGACTTTCGCAAGATCAAGAGTTGCTG 1374  
QY 484 GACGGTTTGTATTATCGCGGGGTGCGCGCGCAATACCGCGACACCGGACGGTGGATT 543  
DB 1373 GAGGAGTTCTGTGTTACTTTGAGACGCGCATGCCCTACTCGCATCTGGCGCATGAACATC 1314  
QY 544 TTCGCGCTGGCGGTTCGCGGCAAGCTGATCTGTTCCGCTGGGTGGTTTCGGCGCA 603  
DB 1313 CCGGACTTGAACCGGACGCGTCCGCGCGCGCGCGGGTGGTCCCGCGACCAAC 1254  
QY 604 GCAGCATTTGCACGCCCGCTGTCCAGCCCCAAGGTGTGGCGCTGATCAACGTGCTGTG 663  
DB 1253 AAGTCCATCTTCTTCAGCGAGCCCTACAGTACTTCTCGCACGCCAGCGTTTATCTTC 1194  
QY 664 CGAGTTGTGATGACCGCATTTGGCCATCAAACTGATTTGA 703  
DB 1193 AACATCGTGATGTCTTCGACACCATCATGAAGCGCAGA 1154

RESULT 14  
US-09-925-300-760/c  
Sequence 760, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 760  
LENGTH: 2048  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1957)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (1963)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc feature  
LOCATION: (2006)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-300-760

Query Match 4.8%; Score 34; DB 10; Length 2048;  
Best Local Similarity 56.1%; Pred. No. 0.64;  
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 393 GGTGGAGTGCAGCTGATTAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGCAATCGT 452  
DB 1381 GGTGACCACGGTCAATGACATGTTGGCGGTGCTGCAGCCCTCGCCGCTGCTGGGAGCCT 1322  
QY 453 GCTGACCTGGTTGAACCCGATCGTATTGGACGCGGTTTGTGTTTATCGGCGG 506  
DB 1321 GCTGACCTGGAGGACCGGAAGTGCATTTGACGCGTGTGTTTGTGACAGGCAG 1268  
RESULT 15  
US-09-745-763-198/c  
Sequence 198, Application US/09745763  
Patent No. US20020065394A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
McCoy, John M.  
LaVallie, Edward R.  
Collins-Racie, Lisa A.  
Evans, Cheryl  
Merberg, David  
Treacy, Maurice  
Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
NUMBER OF SEQUENCES: 219  
CORRESPONDENCE ADDRESS:  
ADDRESS: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/745,763  
FILING DATE: 18-Jun-2000  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 198:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2930 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 198:  
US-09-745-763-198

Query Match 4.8%; Score 33.8; DB 10; Length 2930;  
Best Local Similarity 49.7%; Pred. No. 0.91;  
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;  
QY 415 CAGCGGGTTTGGTAAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTGAACCCGAAT 474  
DB 1652 CGGGGGGCTTGGTCCCACTGGGGCCCTGAGGCGCGGCTTCCACAGGGAACACCGGA 1593  
QY 475 GCGTATTGACGCGTTTGTGTTTATCGCGCGGCTGCGCGCAATACCGCGACACCGGA 534  
DB 1592 GCCTTTGGGGCCCTGGGAGCCTTTAGATCTTTTGGCGCCACGCTCTCCGGGGGACCGC 1533





Result No.	Score	Query			ID	Description
		Match	Length	DB		
c	1	57	8.0	15239	1	US-08-390-878-17
	2	57	8.0	4403765	4	US-09-103-840A-2
	3	57	8.0	4411529	4	US-09-103-840A-1
c	4	37.6	5.3	4403765	4	US-09-103-840A-2
c	5	37.6	5.3	4411529	4	US-09-103-840A-1
	6	34.6	4.9	6611	1	US-08-402-282-2
	7	34.6	4.9	6611	1	US-08-508-004-2
	8	34.6	4.9	6611	1	US-08-402-066-2
	9	34.6	4.9	6611	1	US-08-402-068-2
	10	33.6	4.7	1681	4	US-09-434-288-7
	11	33	4.6	4826	4	US-09-192-983-3
	12	32.8	4.6	3955	4	US-09-214-378-4
	13	32.8	4.6	4315	3	US-08-882-046-3
	14	32.8	4.6	4483	3	US-08-611-729A-7
c	15	32.8	4.6	4508	5	PCT-US93-06251-34
	16	32.4	4.6	707	4	US-08-998-416-940
	17	32.4	4.6	4689	4	US-09-105-537-34
	18	32.4	4.6	36778	4	US-09-105-537-5
	19	32.4	4.6	38506	3	US-09-320-878-19
	20	32.2	4.5	1404	4	US-08-979-608A-10
	21	32	4.5	43280	2	US-08-804-227C-1
	22	31.8	4.5	1050	4	US-09-161-241-77
c	23	31.8	4.5	4550	4	US-09-338-907-182
c	24	31.8	4.5	4550	4	US-09-218-207-182
	25	31.6	4.4	50937	4	US-09-428-517-1
c	26	31.4	4.4	50937	4	US-09-428-517-1
c	27	31.2	4.4	219	3	US-08-078-271B-9





```

Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 359 CGCGGTGGCCACTGACACGGCAACCGGTGGGGTGGAGGTGAGCGTCAATAGCAGC 418
      |||||
Db 619 CGCGCATCGCTAAGCGCTTCCGAAGCGGTGGCGGTGAAGTTTACGCACGCGGTGCGC 678
      |||||
QY 419 GGGTTTGGTTAAAGCCCATCTTGTATGCAATCGTGTGACCTGTTGCAACCGGATGCGT 478
      |||||
Db 679 GGGCTGTGGTGCAGCGGGGGCGCGGAGTTTACGCTGCCACGATGGACCCCGGGAGT 738
      |||||
QY 479 ATTGACGCGCTTGTGTTTATCGCGCGCGTGGCGGCAATACGGCGA 527
      |||||
Db 739 TCCGCACTGCCGCGCTGCCACCGAGGGGCGATCGTGGACGGCA 787
      |||||

```

## RESULT 7

```

US-08-508-004-2
; Sequence 2, Application US/08508004
; Patent No. 5582969
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/508,004
; FILING DATE: 27-JUL-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/402,282
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6611 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 390..1538
; OTHER INFORMATION: /function= "coding sequence"
; OTHER INFORMATION: /product= "DNA polymerase"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2107..3132
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: /frame= "frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3138..3359
; OTHER INFORMATION: /function= "potential open reading"

```

```

; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 4690..5028
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5028..5375
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5375..5653
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 5653..5910
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 6078..6491
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: frame"
; US-08-508-004-2
; Query Match 4.9%; Score 34.6; DB 1; Length 6611;
; Best Local Similarity 50.3%; Pred. No. 1.7;
; Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
QY 359 CGCGGTGGCCACTGACACGGCAACCGGTGGGGTGGAGGTGAGCGTCAATAGCAGC 418
      |||||
Db 619 CGCGCATCGCTAAGCGCTTCCGAAGCGGTGGCGGTGAAGTTTACGCACGCGGTGCGC 678
      |||||
QY 419 GGGTTTGGTTAAAGCCCATCTTGTATGCAATCGTGTGACCTGTTGCAACCGGATGCGT 478
      |||||
Db 679 GGGCTGTGGTGCAGCGGGGGCGCGGAGTTTACGCTGCCACGATGGACCCCGGGAGT 738
      |||||
QY 479 ATTGACGCGCTTGTGTTTATCGCGCGCGTGGCGGCAATACGGCGA 527
      |||||
Db 739 TCCGCACTGCCGCGCTGCCACCGAGGGGCGATCGTGGACGGCA 787
      |||||
RESULT 8
US-08-402-066-2
; Sequence 2, Application US/08402066
; Patent No. 5612182
; GENERAL INFORMATION:
; APPLICANT: Pearson, Robert E.
; APPLICANT: Dickson, Julie A.
; APPLICANT: Hamilton, Paul T.
; APPLICANT: Little, Michael C.
; APPLICANT: Beyer Jr., Wayne F.
; TITLE OF INVENTION: MYCOBACTERIOPHAGE SPECIFIC FOR THE
; TITLE OF INVENTION: MYCOBACTERIUM TUBERCULOSIS COMPLEX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/402,066
; FILING DATE: 10-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3283
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6611 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 390..1538
; OTHER INFORMATION: /function= "coding sequence"
; OTHER INFORMATION: /product= "DNA polymerase"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2107..3132
; OTHER INFORMATION: /function= "potential open reading"
; OTHER INFORMATION: /frame= "frame"
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3138..3359
; OTHER INFORMATION: /function= "potential open reading"

```



; OTHER INFORMATION: /function= "potential open reading  
; OTHER INFORMATION: frame"  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 6078..6491  
; OTHER INFORMATION: /function= "potential open reading  
; OTHER INFORMATION: frame"  
US-08-402-068-2

Query Match 4.9%; Score 34.6; DB 1; Length 6611;  
Best Local Similarity 50.3%; Pred. No. 1.7;  
Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;  
QY 359 CGCGGTGGGCACTGACACCGCAACCGGTGGGGTGGAGGTGAGGTGCTGATAAGCAGC 418  
DB 619 CGCGCATCGCTAAGGGCTTCCCAAGCGGTGCCGGTGAAGTTTACGACAGCGGTGCGC 678  
QY 419 GGGTTTGGGTAAGCCCATCTGTGATGCAATCTGTGACCTGTGCTGCTTGAACCCGAATGCGT 478  
DB 679 GGCTGTGTGACAGGGGGGGCGGGAGTTTACGCTGCCACGATGGACCCCGGGAGT 738  
QY 479 ATTTGACGCGTTTGTGTTTATCGCGGGCGTGGCGGCAATACGGCGA 527  
DB 739 TCCGCAACTGCCCGGCTGCCACCGAGCGGGCATCTGTGACGGCGA 787

RESULT 10  
US-09-434-288-7  
; Sequence 7, Application US/09434288  
; Patent No. 6303767  
; GENERAL INFORMATION:  
; APPLICANT: Betlach C., Melanie  
; APPLICANT: McDaniel, Robert  
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
; FILE REFERENCE: 30062-20030.00  
; CURRENT APPLICATION NUMBER: US/09/434,288  
; CURRENT FILING DATE: 1999-11-05  
; PRIOR APPLICATION NUMBER: 60/107,093  
; PRIOR FILING DATE: 1998-11-05  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 1681  
; TYPE: DNA  
; ORGANISM: Streptomyces narbonensis  
US-09-434-288-7

Query Match 4.7%; Score 33.6; DB 4; Length 1681;  
Best Local Similarity 49.4%; Pred. No. 1.8;  
Matches 87; Conservative 0; Mismatches 89; Indels 0; Gaps 0;  
QY 520 TACGGGACACCGGACCGGTGATTTTCGCGCTGGCGGTTCGGCGCAAGCCTGATCTGG 579  
DB 1493 TACTGCGGTCCGGCGGTTCCTCCACGACGGCGGAGTTCGACGCGGACTTCTTCGG 1552  
QY 580 TTCCCGTGTGGTTTTCGGCGGAGCAGATTTTCACCGCGGTGTCCAGCCCCAAGGTG 639  
DB 1553 ATCTGCGCGCGGAGGCGCTCGCCATGGACCGCGAGCGGGGTGTCCCTCACCACCGCG 1612  
QY 640 TGGCGCTGGATCAACGTCGTGTGGAGTTGTGATGACCGCATTCGGCCATCAACT 695  
DB 1613 TGGAGGCGATCGAGCAGCGGGGATCGACCCGACGAGCTGAAGGGCAGCGGCT 1668

RESULT 11  
US-09-192-983-3  
; Sequence 3, Application US/09192983A  
; Patent No. 6242244  
; GENERAL INFORMATION:  
; APPLICANT: Donohue, Timothy  
; APPLICANT: Barber, Robert  
; APPLICANT: Witthuhn, Vernon

; TITLE OF INVENTION: Microbial System for Formaldehyde Sensing and  
; TITLE OF INVENTION: Remediation  
; FILE REFERENCE: 960296.95505  
; CURRENT APPLICATION NUMBER: US/09/192,983A  
; CURRENT FILING DATE: 1998-11-16  
; EARLIER APPLICATION NUMBER: 08/919,953  
; EARLIER FILING DATE: 1997-08-29  
; EARLIER APPLICATION NUMBER: 08/608,241  
; EARLIER FILING DATE: 1996-02-28  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 4826  
; TYPE: DNA  
; ORGANISM: Rhodobacter sphaeroides  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (215)..(895)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (993)..(2165)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2236)..(4437)  
US-09-192-983-3

Query Match 4.6%; Score 33; DB 4; Length 4826;  
Best Local Similarity 46.7%; Pred. No. 4.4;  
Matches 105; Conservative 0; Mismatches 120; Indels 0; Gaps 0;  
QY 288 CAAGTGGNAGCGCCACAGATCATTTGAAGAAACAGAACCAACCGTCCCGATGACACGCC 347  
DB 3990 CATGTGCAAGTGGCCCGCGCGAAAGTGGCGCGCTCGTCCGCGCATGCTCGGC 4049  
QY 348 TTTGGCGGTTTCGGCGGTGGCCACTGACACGCGCAACCGGGTGGAGGTGAGCGT 407  
DB 4050 GGTGGCGGCGAGCCAGGTGGCGGTGCAGGACAGATCGCTTCTCTGTCGAGATGA 4109  
QY 408 CGATAAGCAGCGGTTTGGGTAAAGCCCATTTGATGGAATCGTGTGACCTGTTGAA 467  
DB 4110 CGACGACCTGCTGAGCGCATGGGACTGCTGCTGGAGAAATGGGGGTGAGCGTGTCTGA 4169  
QY 468 CCGCATGCTATTGGACGCGTTTGTGTTTATCGCGCGGTCGG 512  
DB 4170 CGCGCCTCGGGCGAGGCGCTGCGCTGATCGAGGAGATCGG 4214

RESULT 12  
US-09-214-278-4  
; Sequence 4, Application US/09214278  
; Patent No. 6291210  
; GENERAL INFORMATION:  
; APPLICANT: Sakano, Seiji  
; APPLICANT: Itoh, Akira  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8576  
; CURRENT APPLICATION NUMBER: US/09/214,278  
; CURRENT FILING DATE: 1999-01-26  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 3955  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (12)..(3725)  
; NAME/KEY: sig\_peptide  
; LOCATION: (12)..(89)  
; NAME/KEY: mat\_peptide  
; LOCATION: (90)..(3725)  
US-09-214-278-4

Query Match 4.6%; Score 32.8; DB 4; Length 3955;  
Best Local Similarity 58.0%; Pred. No. 4.6;  
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 378 GCGCAACCGGTGGGGTGGAGGTGCGATGAGCAGCGGGTTGGGTAAGCCCAT 437  
Db 3353 GAGGAGCGGTGCGCGGGGAGGAGCGCAACACAGTGGCGCCGCTCAACCCCAT 3412  
QY 438 GTTGATGCAATCTGCTGACCTGCTTCAACCCGAATGCG 477  
Db 3413 CGCAACCCCATGAGCGCGCGGGGGGCCACAGGACGTG 3452

## RESULT 13

US-08-882-046-3  
; Sequence 3, Application US/08882046  
; Patent No. 6136952

## GENERAL INFORMATION:

APPLICANT: Li, Linheng  
APPLICANT: Hood, Leroy  
APPLICANT: Krantz, Ian D.  
APPLICANT: Spinner, Nancy B.  
TITLE OF INVENTION: Human Jagged Polypeptide, Encoding  
TITLE OF INVENTION: Nucleic Acids and Methods of Use  
NUMBER OF SEQUENCES: 110  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/882,046  
FILING DATE: 25-JUN-1997  
CLASSIFICATION: 536

## ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UW 2637  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949

## INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4315 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:

NAME/KEY: CDS  
LOCATION: 16..3460

US-08-882-046-3

Query Match 4.6%; Score 32.8; DB 3; Length 4315;  
Best Local Similarity 58.0%; Pred. No. 4.8;  
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 378 GCGCAACCGGTGGGGTGGAGGTGAGCGTGAATGAGCAGCGGGTTGGGTAAGCCCAT 437  
Db 3090 GAGGAGCGGTGCGCGGGGAGGAGCGCAACACAGTGGCGCCGCTCAACCCCAT 3149

QY 438 GTTGATGCAATCTGCTGACCTGCTTGAACCCGAATGCG 477

Db 3150 CGCAACCCCATGAGCGCGCGGGGGGCCACAGGACGTG 3189

## RESULT 14

US-08-611-729A-7

; Sequence 7, Application US/08611729A  
; Patent No. 6004924

## GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
APPLICANT: Gray, Grace E.

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611,729A  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Misrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE

## INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 332..4102

US-08-611-729A-7

Query Match 4.6%; Score 32.8; DB 3; Length 4483;  
Best Local Similarity 58.0%; Pred. No. 4.9;  
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 378 GCGCAACCGGTGGGGTGGAGGTGAGCGTGAATGAGCAGCGGGTTGGGTAAGCCCAT 437  
Db 3163 GAGGAGCGGTGCGCGGGGAGGAGCGCCAAACAGTGGCGCCGCTCAACCCCAT 3222

QY 438 GTTGATGCAATCTGCTGACCTGCTTGAACCCGAATGCG 477

Db 3223 CGCAACCCCATGAGCGCGCGGGGGGCCACAGGACGTG 3262

## RESULT 15

PCT-US93-06251-34/c

; Sequence 34, Application PC/TUS9306251

## GENERAL INFORMATION:

APPLICANT: Wickstrom, Eric and Rife, Jason P.  
TITLE OF INVENTION: Trivalent Synthesis of Oligonucleotides Containing  
TITLE OF INVENTION: Stereospecific Alkylphosphonates and Arylphosphonates  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:



ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER  
STREET: 400 Garden City Plaza  
CITY: Garden City  
STATE: NY  
COUNTRY: USA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/06251  
FILING DATE: 19930630  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Digiglio, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 8586  
TELEPHONE: 516-742-4343  
TELEFAX: 516-742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 34:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4508 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
PCT-US93-06251-34

Query Match 4.6%; Score 32.8; DB 5; Length 4508;  
Best Local Similarity 48.9%; Pred. No. 4.9;  
Matches 88; Conservative 0; Mismatches 92; Indels 0; Gaps 0;  
Qy 412 AAGCAGCGGGTTGGGTAAGCCCATGTTGATGGCAATCGTGTGACCTGGTTGAACCCG 471  
Db 2366 AGGAGCCGTAATTTGGCGTACTCCACGATGAGGAGCGGGCCATCTCGCTGACGCC 2307  
Qy 472 AATCGTATTTGGACGGCGTTTGTGTTTATCGCGCGCGTCCGCGCAATACGGCGACCC 531  
Db 2306 CCATACAATTTGATGACATGTGGTGTGTTGACCTGCTTCAGGACGTTGAACCTGTGACAGC 2247  
Qy 532 GGACGGTGGATTTTCGCCGCTGGCGGCTTCGGCGGAGCCTGATCTGGTTCCCGCTGGTG 591  
Db 2246 AGGTCTGAAGCTCACTCGGGGAGCGTTCTCTTTTCAAGCATCTTCAGGCCACCGTGTG 2187

Search completed: April 27, 2003, 03:47:50  
Job time : 10403.1 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 18:49:18 ; Search time 2249.77 Seconds  
(without alignments)  
11293.030 Million cell updates/sec

Title: US-09-105-117K-1\_COPY\_1421\_2293  
Perfect score: 873  
Sequence: 1 gtcgataagcagcggtttg.....ctgtgagctctgaccgtag 873

Scoring table: IDENTITY NUC  
Gapop 10\_0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	873	100.0	2374	1	CGLYSEG	X96471 C.glutamici
2	873	100.0	2374	6	A93933	AP005277 Coryneb
C 3	873	100.0	333150	1	AP005277	AX127147 Sequence
C 4	873	100.0	349980	6	AX127147	AX123538 Sequence
C 5	568	65.1	627	6	AX123538	AX067087 Sequence
C 6	404.2	46.3	993	6	AX067087	AX063771 Sequence
C 7	326	37.3	822	6	AX063771	AX244059 Sequence
C 8	326	37.3	822	6	AX244059	AX123539 Sequence
C 9	303	34.7	708	6	AX123539	AB083133 Coryneb
C 10	193.6	22.2	1771	1	AB083133	AP003001 Mesorhiz
C 11	141.2	16.2	345783	1	AP003001	AE008076 Agrobacte
C 12	136.6	15.6	8472	1	AE008076	AE009111 Agrobacte
C 13	136.6	15.6	13051	1	AE009111	AL450223 Streptomy
C 14	133.6	15.3	15586	1	SC7H9	AL591788 Sinorhizo
C 15	115	13.2	306250	1	SME591788	AE011965 Xanthomon
C 16	88.6	10.1	10871	1	AE011965	AL049863 Streptomy
C 17	87.2	10.1	36583	1	SC5H1	M37389 Pseudomonas
C 18	86.2	10.0	1525	1	PSEIGRA	AE002063 Deinococc
C 19	86.2	9.9	12257	1	AE002063	AE005530 Escherich
C 20	84.8	9.7	10336	1	AE005530	AP002563 Escherich
C 21	83.6	9.6	266658	1	AP002563	U28377 Escherichia
C 22	80.4	9.2	11024	1	AE000382	AC020884 Mus muscu
C 23	80.4	9.2	141744	1	ECU28377	AL627274 Salmonell
C 24	79.4	9.1	110000	2	AC020884	AE008808 Salmonell
C 25	76.4	8.8	256050	1	AL627274	AL646059 Ralstonia
C 26	74.8	8.6	22929	1	AE008808	AR012071 Sequence
C 27	73.4	8.4	190050	1	AL646059	AR025196 Sequence
C 28	71.6	8.2	5541	6	AR012071	AR038653 Sequence
C 29	71.6	8.2	5541	6	AR025196	AR068500 Sequence
C 30	71.6	8.2	5541	6	AR038653	AX343066 Sequence
C 31	71.6	8.2	5541	6	AR068500	AX347518 Sequence
C 32	71.6	8.2	5541	6	AX343066	AX347564 Sequence
C 33	71.6	8.2	5541	6	AX347518	AL357613 Streptomy
C 34	71.6	8.2	5541	6	AX347564	AX331662 Sequence
C 35	71.6	8.2	5541	6	SC5F8	AE004852 Pseudomon
C 36	65.6	7.5	33285	1	SC5F8	U65741 Aeromonas s
C 37	65.4	7.5	988	6	AX431662	AE012410 Xanthomon
C 38	63.4	7.3	10542	1	AE004852	U34849 Mycobacteri
C 39	61.6	7.1	4033	1	ASU65741	AE007056 Mycobacte
C 40	60	6.9	213050	1	AL646067	I86263 Sequence 17
C 41	59.6	6.8	10747	1	AE012410	Z74025 Mycobacteri
C 42	57	6.5	14844	1	MEU34849	
C 43	57	6.5	14869	1	AE007056	
C 44	57	6.5	15239	6	I86263	
C 45	57	6.5	38500	1	MTCY39	

ALIGNMENTS

RESULT 1  
CGLYSEG

LOCUS C.glutamici 2374 bp DNA linear BCT 18-MAR-2001  
DEFINITION C.glutamici lysE and lysG genes.  
ACCESSION X96471

VERSION X96471.1 GI:1729753  
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine

SOURCE Corynebacterium glutamicum.  
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

REFERENCE 1 (bases 1 to 2374)  
AUTHORS Vrijic,M., Sahm,H. and Eggeling,L.

TITLE A new type of transporter with a new type of cellular function:  
 L-lysine export from *Corynebacterium glutamicum*  
 JOURNAL Mol. Microbiol. 22 (5), 815-826 (1996)  
 MEDLINE 97126810  
 PUBMED 8971704  
 REFERENCE 2 (bases 1 to 2374)  
 AUTHORS Vrljic, M.M.  
 DIRECT SUBMISSION  
 TITLE Submitted (07-MAR-1996) M.M. Vrljic, Institut fuer Biotechnologie  
 JOURNAL 1, Forschungszentrum Juelich, Postfach 1913, D-52425 Juelich, FRG  
 FEATURES Location/Qualifiers  
 1..2374  
 /organism="Corynebacterium glutamicum"  
 /strain="R127"  
 /db\_xref="taxon:1718"  
 complement (82..954)  
 /gene="lysG"  
 complement (82..954)  
 /gene="lysG"  
 /note="Lysine governor"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Lysine export regulator protein"  
 /protein\_id="CAA65323.1"  
 /db\_xref="GI:1729754"  
 /db\_xref="SWISS-PROT:P94632"  
 /translation="MNPIDLTLLSIIDEGSEFASLALSI8PSAVSRVKALEHHVG  
 RVLVSTPAKATEAGEVLAQAKLVLLQAEKQALSLAEPLITIAINADSLSTW  
 RFLVFNVAWSGATLTLEDEAMPTLSLRGDLVLAETREANPVAGCEVVELGTMR  
 HLAATPSLRDAVMVDKLDWAMPVLRFGKVDLQDRDLDRVDPGVRERVSIVPS  
 ASGFGRIRGLGWGLLPETQAPMLKAGEVILLDEIPDTPMYWRWRLESRLARL  
 TRAVDAALEGGRP"  
 1025..1726  
 /gene="lysE"  
 1025..1726  
 /gene="lysE"  
 /codon\_start=1  
 /transl\_table=11  
 /product="Lysine exporter protein"  
 /protein\_id="CAA65324.2"  
 /db\_xref="GI:13397387"  
 /db\_xref="SWISS-PROT:P94633"  
 /translation="MEIFITGLLGSALLSIIGPQNVIVKIGIKREGILVILVCLL  
 SDVFLIAGTLDLSNAPIVLDIMRWGGIAYLLWFAVMAKADMTINKVAPQILE  
 ETEPTVDDPLGSAVDATDRNRVSVSDKQVVRVPLMAIVLWLNPNAYLDA  
 FVPIGVGAQYDGTGRWIFPAAGAFASLIWFFLVGFAGAAALSRPLSSPKVWRWVNVV  
 AVMTALAIKLMNG"  
 complement (1723..2352)  
 /gene="orf3"  
 complement (1723..2352)  
 /gene="orf3"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAA65325.1"  
 /db\_xref="GI:1729756"  
 /db\_xref="SPRMBEL:P94634"  
 /translation="WVALRDIVASGKALYVIGSYGPFLTAFAAEPMEEGFLLIHQ  
 PYSIINRWEEPCDDGNLQSLAANNGLVAFSPAQGLLTKYLDIGIPGSRASQ  
 GKSLSGMLNVRKINDIAQERGSQAOMALAWVLRQEGYADTVTSALIGA  
 SSVEQLDNLSDLSLNNLEFSDAKLEATDEISHDAGINIMAKATDSKTREN"

## BASE COUNT

526 a

640 c

648 g

560 t

## ORIGIN

Query Match 100.0%; Score 873; DB 1; Length 2374;  
 Best Local Similarity 100.0%; Pred. No. 4,7e-185;  
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCGATAAGCAGCGGTTTCGGTAAGCCCATGTTGATGCAATCGTGCACCTGGTGG 60  
 DB 1421 GTCGATAAGCAGCGGTTTCGGTAAGCCCATGTTGATGCAATCGTGCACCTGGTGG 1480  
 QY 61 AACCCGAATCGTATTGGACGGTTTGTGTTTATCGCGCGCGTGCAGCGCAATACGGC 120  
 1481 AACCCGAATCGTATTGGACGGTTTGTGTTTATCGCGCGCGTGCAGCGCAATACGGC 1540  
 QY 121 GACACCGAGCGTGGATTTTCGCCGCTGGCGGTTCGCGGCAAGCCCTGATCTGGTTCCCG 180  
 DB 1541 GACACCGAGCGTGGATTTTCGCCGCTGGCGGTTCGCGGCAAGCCCTGATCTGGTTCCCG 1600  
 QY 181 CTGGTGGGTTTCGGCGCAGCAGATTGTCAAGCCGCTGTCCAGCCCAAGAGTGTGGCGC 240  
 DB 1601 CTGGTGGGTTTCGGCGCAGCAGATTGTCAAGCCGCTGTCCAGCCCAAGAGTGTGGCGC 1660  
 QY 241 TGGATCAACGTCGTCTGTCGAGTTGTGATGACGCGATTGCCCATCAAACTGATGTTGATG 300  
 DB 1661 TGGATCAACGTCGTCTGTCGAGTTGTGATGACGCGATTGCCCATCAAACTGATGTTGATG 1720  
 QY 301 GGTAGTTCGCGGCTTTTGGAAATCGGTGGCTTCGCCCAAAATGTTGATGCGCGGCTCG 360  
 DB 1721 GGTAGTTCGCGGCTTTTGGAAATCGGTGGCTTCGCCCAAAATGTTGATGCGCGGCTCG 1780  
 QY 361 TGGGAAATCTCATCGATCGCTCCAACTCGGCGTCAGAAAACCTCCAAAGTGTGATGTA 420  
 DB 1781 TGGGAAATCTCATCGATCGCTCCAACTCGGCGTCAGAAAACCTCCAAAGTGTGATGTA 1840  
 QY 421 TCAAGGCTGTGTCCAGCTGCTCAACTGACGAGCAACCAATCAATGCACTGGTTCAGGTA 480  
 DB 1841 TCAAGGCTGTGTCCAGCTGCTCAACTGACGAGCAACCAATCAATGCACTGGTTCAGGTA 1900  
 QY 481 TCGCGCCGCTACTCTCTTCTCGCGCAGCAGCCCATGCAAGCGCCCATCTGCGCAAGTGAC 540  
 DB 1901 TCGCGCCGCTACTCTCTTCTCGCGCAGCAGCCCATGCAAGCGCCCATCTGCGCAAGTGAC 1960  
 QY 541 TCGCGCGGTTCTCGGCGGATGTCATGAGCTTGGCGGACCATATCAATATGTTGACGTTTC 600  
 DB 1961 TCGCGCGGTTCTCGGCGGATGTCATGAGCTTGGCGGACCATATCAATATGTTGACGTTTC 2020  
 QY 601 AACATGCCCTCAGACAGGACTTACCCTGGCTGGCGCGGAAACCTCTCGAATTCATCG 660  
 DB 2021 AACATGCCCTCAGACAGGACTTACCCTGGCTGGCGCGGAAACCTCTCGAATTCATCG 2080  
 QY 661 AGATATTTGTCGTGACGAGGCGCTCGCGCAAGTGTGAGAAAGCAATGACGCCAAGACCA 720  
 DB 2081 AGATATTTGTCGTGACGAGGCGCTCGCGCAAGTGTGAGAAAGCAATGACGCCAAGACCA 2140  
 QY 721 TTGTTGGCAGCTGACGCAACAAAGTTCTCACCGTCATCGCCCGTTCCTCCACCAACGA 780  
 DB 2141 TTGTTGGCAGCTGACGCAACAAAGTTCTCACCGTCATCGCCCGTTCCTCCACCAACGA 2200  
 QY 781 TTAATCATGGAATAGCTTCGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCCATGAAC 840  
 DB 2201 TTAATCATGGAATAGCTTCGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCCATGAAC 2260  
 QY 841 TCAGCCGCTCCGCTGTGAGCTCTGGACCGTAG 873  
 DB 2261 TCAGCCGCTCCGCTGTGAGCTCTGGACCGTAG 2293

## RESULT 2

A93933

LOCUS

DEFINITION

A93933

ACCESSION

VERSION

A93933.1

GI:6742037

KEYWORDS

SOURCE

ORGANISM

Corynebacterium glutamicum.

Corynebacterium glutamicum

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Corynebacteriaceae;

Corynebacterium.

1 (bases 1 to 2374)

Vrljic, M. and Eggeling, L.

PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED

ACTIVITY OF EXPORT CARRIERS

Patent: WO 9723597-A 2 03-JUL-1997;

KERNFORSCHUNGSANLAGE JUELICH (DE); VRLIJC MARINA (DE)

FEATURES		Location/Qualifiers	
source		1. .2374	
		/organism="Corynebacterium glutamicum"	
		/db_xref="taxon:1718"	
BASE COUNT	526 a	540 c	560 t
ORIGIN			
Query Match 100.0%; Score 873; DB 6; Length 2374;			
Best Local Similarity 100.0%; Pred. No. 4.7e-185;			
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	GTCCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGGCAATCGCTGACCTGGTTG	60
Db	1421	GTCCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGGCAATCGCTGACCTGGTTG	1480
QY	61	AACCCGAATGCGTATTTCGACGCGTTTGTGTTTATCGCGCGCTCGCGCGCAATACGGC	120
Db	1481	AACCCGAATGCGTATTTCGACGCGTTTGTGTTTATCGCGCGCTCGCGCGCAATACGGC	1540
QY	121	GACACCGGACGCTGGATTTTCGCCCTGCGCGCTTTCGCGCAAGCCTGATCTGCTCCG	180
Db	1541	GACACCGGACGCTGGATTTTCGCCCTGCGCGCTTTCGCGCAAGCCTGATCTGCTCCG	1600
QY	181	CTGCTGGGTTTTCGCGGACGAGCATTTGTCAGCCGCTGTCAGCCCAAGGTGTGGCG	240
Db	1601	CTGCTGGGTTTTCGCGGACGAGCATTTGTCAGCCGCTGTCAGCCCAAGGTGTGGCG	1660
QY	241	TGGATCAACGTCGTCGTCGCGATGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG	300
Db	1661	TGGATCAACGTCGTCGTCGCGATGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG	1720
QY	301	GGTTAGTTTTCGCGGTTTGGAAATCGGTGCGCTTCGCCCAAAATGTTGATCGCGCGCTCG	360
Db	1721	GGTTAGTTTTCGCGGTTTGGAAATCGGTGCGCTTCGCCCAAAATGTTGATCGCGCGCTCG	1780
QY	361	TGGGAATCTCATGATCGCTCCAACTCGCGGTGAGAACTCCAAGTTGTTGAGTGA	420
Db	1781	TGGGAATCTCATGATCGCTCCAACTCGCGGTGAGAACTCCAAGTTGTTGAGTGA	1840
QY	421	TCAAGGCTGTTGTCAGCTCTCAACTGACGAAGCACCAATCAATGACCTGCTCAGCGTA	480
Db	1841	TCAAGGCTGTTGTCAGCTCTCAACTGACGAAGCACCAATCAATGACCTGCTCAGCGTA	1900
QY	481	TCGCGCGCTACTCTCTTCTGCTCGCGGACCCATGCAAGCGCCATCTGCGCAAGTGAC	540
Db	1901	TCGCGCGCTACTCTCTTCTGCTCGCGGACCCATGCAAGCGCCATCTGCGCAAGTGAC	1960
QY	541	TGCGCGGTTCTGCGGCGATGTCATTGAGCTTGGCGACCATATCAATTTGTTACGTTT	600
Db	1961	TGCGCGGTTCTGCGGCGATGTCATTGAGCTTGGCGACCATATCAATTTGTTACGTTT	2020
QY	601	AACATGCCCTCAGACAGGACTTACCTGCTGCGCGGGAACCTCTGGAATTTCCATCG	660
Db	2021	AACATGCCCTCAGACAGGACTTACCTGCTGCGCGGGAACCTCTGGAATTTCCATCG	2080
QY	661	AGATATTGTCGTCGACGAGCCCTGCGCAAGTGTGAGAAAGCAATGAGCCCAAGACCA	720
Db	2081	AGATATTGTCGTCGACGAGCCCTGCGCAAGTGTGAGAAAGCAATGAGCCCAAGACCA	2140
QY	721	TTGTTGGCAGTCACTTGCAACAAGTTCTACCGTTCATGCGCCCGGTTCTCCACCAACGA	780
Db	2141	TTGTTGGCAGTCACTTGCAACAAGTTCTACCGTTCATGCGCCCGGTTCTCCACCAACGA	2200
QY	781	TTAATGATGGAATAGCTTGCTGATGATGATCAGAGCGGCGAGCCCTCTCGGCATGAC	840
Db	2201	TTAATGATGGAATAGCTTGCTGATGATGATCAGAGCGGCGAGCCCTCTCGGCATGAC	2260
QY	841	TCAGCCGCTCCGCTGTGAGCTCTGGACCGTAG	873
Db	2261	TCAGCCGCTCCGCTGTGAGCTCTGGACCGTAG	2293

RESULT 3

RESULT 3

AP005277/c	AP005277	333150 bp	DNA	linear	BCT 08-AUG-2002
LOCUS	Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 4/10.				
DEFINITION					
ACCESSION	AP005277	BA000036			
VERSION	AP005277.1	GI:21323710			
KEYWORDS	Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.				
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.				
ORGANISM					
REFERENCE	1 Nakagawa,S.				
AUTHORS	Complete genomic sequence of Corynebacterium glutamicum ATCC 13032				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 333150)				
REFERENCE	Nakagawa,S.				
AUTHORS	Direct Submission				
TITLE	Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co.				
JOURNAL	Ltd., Tokyo Research Laboratories; 3-6-6, Asehi-machi, Machida, Tokyo 194-8533, Japan (E-mail:snakagawa@xanagen.com, Tel:81-44-829-3031, Fax:81-44-813-1651)				
COMMENT	This sequence is conducted by collaboration of Kyowa Hakko Kogyo Co. Ltd. And Kitasato University.				
FEATURES	Location/Qualifiers				
source	1. .333150				
	/organism="Corynebacterium glutamicum ATCC 13032"				
	/strain="ATCC 13032"				
	/db_xref="taxon:196627"				
gene	/note="ATCC 13032"				
	complement(63..1292)				
CDS	/gene="Cg10944"				
	complement(63..1292)				
	/gene="Cg10944"				
	/codon_start=1				
	/transl_table=11				
	/product="Hypothetical membrane protein"				
	/protein_id="BAB98337.1"				
	/db_xref="GI:21323711"				
	/translation="MTNPTERNARLLIWANLQNIQDQIVAAKTLPWLLQAGAPG				
	FLALLVPIRAGSMPLPQAIATGWLRQTSKVMVIGNSQGVFVSALGIGVAALFLRG				
	WAGLTVIYLLAALSFRSMGSIASKDVQKVIKSGKGLVTRATVIGVMGLVAGL				
	AIAIPLGSHSPTRVLAIVIASSPSLFAVIFARIYAKPATKPNASPNVWRRCI				
	AALKDKAPRRFLVRSMVLTAFTAFIVALAESGNSIDSLGPFLLASGLASVMVG				
	RISGWDHSSKNWAGALPGSIVLILVLSAPAPQINTLPPLSFLLTLAHTA				
	IRVARKTYMDMAEGDQTRYVADANTLMGVLLIVGALSFGFIATFGNEAALLFLAAI				
	GLLGTISARGLKEVSAG"				
gene	1382..2917				
CDS	/gene="Cg10945"				
	1382..2917				
	/gene="Cg10945"				
	/codon_start=1				
	/transl_table=11				
	/product="Putative multicopper oxidases"				
	/protein_id="BAB98338.1"				
	/db_xref="GI:21323712"				
	/translation="MPELNRRTPFKAGVLAATVGAQVLVACSSDDVRGVGGRPTLL				
	PIPADLGTREGSVHFALAEQGESQILDVTTKTTWGFNGHTLGTPLVKKGGDVHV				
	DVINLDEMTHVHMKLPAIADGPHSPICPGQVTSPTWTVANDATLWHPHTHG				
	LTGLHVRGLAGMIIVDEATDKLDLREYGVDDIPLVMDHRLFEDGSLDEDLPL				
	GLIGDTPANGITNAHPDATRRVRFRVLNGSNRFRYNLAESDTRTFQVIASDGLLD				
	EPQRTLIAIGPGRWEIIVLEPEGEDVTLESVFEDNIGVPDFDFVFGMSDSFQL				
	LITGPSDDAAQALPGVLVKSTEPDVIDATETFTMTFESINDLQMDMRQVDVID				
	HDQPEVMIVTNDNSDPHFHVDARFKVLEFGTDELNDGDKDVTGLPPGATATL				
	AVEFGHYDPQPMYHCHMLYHEDQGMWGQFVIVEPGDEPAAVLGSGTSSIDSAGG				
	HAH"				
gene	3063..3983				
CDS	/gene="Cg10946"				
	3063..3983				
	/gene="Cg10946"				
	/note="PF00005:ABC transporter				
	TIGR00960:3a0501s02: Type II (General) Secretory Pathway				
	(IISP) Family protein				



```

Db 328418  CTGGTGGGTTTCGGCGCAGCAGCATGTTCAGCCGCTGTTCAGCCCAAGGTGTGGCG 328359
QY 241  TGGATCAACGTCGTCTGGCAGTGTGTGATGACCGCATGGCCATCAAACTCATGTTGATG 300
Db 328358  TGGATCAACGTCGTCTGGCAGTGTGTGATGACCGCATGGCCATCAAACTCATGTTGATG 328299
QY 301  GGTAGTTTTCCGCGGTTTTGGAATCGGTGGCTTCGCCCAAAATGTTGATGCCGCGCTCG 360
Db 328298  GGTAGTTTTCCGCGGTTTTGGAATCGGTGGCTTCGCCCAAAATGTTGATGCCGCGCTCG 328239
QY 361  TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAAATCTCAAGTTGTTGAGTGAA 420
Db 328238  TGGGAATCTCATGATCGCTCCAACTCGCGGTCAGAAATCTCAAGTTGTTGAGTGAA 328179
QY 421  TCAAGGCTGTGTCCAGCTGTCTCAACTCAAGCAGCACCATAATGCACTGCTCACGGTA 480
Db 328178  TCAAGGCTGTGTCCAGCTGTCTCAACTCAAGCAGCACCATAATGCACTGCTCACGGTA 328119
QY 481  TCCGCGCGCTACTCTCTTGTCTCGCGCAGCACCATAATGCAAGCGCATCTGCGCAAGTGAC 540
Db 328118  TCCGCGCGCTACTCTCTTGTCTCGCGCAGCACCATAATGCAAGCGCATCTGCGCAAGTGAC 328059
QY 541  TGCCCGGCTTCTGGCGATGTCATTGAGCTTGGCGGAGAACCTCTGGAATTTCCATCG 600
Db 328058  TGCCCGGCTTCTGGCGATGTCATTGAGCTTGGCGGAGAACCTCTGGAATTTCCATCG 327999
QY 601  AACATGCGCTCAGACAGGAGCTTACCTTGGCTGGCGGAGAACCTCTGGAATTTCCATCG 660
Db 327998  AACATGCGCTCAGACAGGAGCTTACCTTGGCTGGCGGAGAACCTCTGGAATTTCCATCG 327939
QY 661  AGATATTGTCGCTGAGCAGCCCTTGGCGAGTGTGAGAAAGCAATGAGCCCAAGACCA 720
Db 327938  AGATATTGTCGCTGAGCAGCCCTTGGCGAGTGTGAGAAAGCAATGAGCCCAAGACCA 327879
QY 721  TTGTTGGCAGCTGACTCAACAGTTCTCACCCTGATCGCCCGTCTTCCACCAAGCA 780
Db 327878  TTGTTGGCAGCTGACTCAACAGTTCTCACCCTGATCGCCCGTCTTCCACCAAGCA 327819
QY 781  TTAATGATGGAATAGCTTGGCTGATGATCAGAAAGCGGAGCCCTCTCGCCCATGAC 840
Db 327818  TTAATGATGGAATAGCTTGGCTGATGATCAGAAAGCGGAGCCCTCTCGCCCATGAC 327759
QY 841  TCAGCCGCTCCGCTGTGAGCTCTGGACCGTAG 873
Db 327758  TCAGCCGCTCCGCTGTGAGCTCTGGACCGTAG 327726

```

```

RESULT 4
AX127147/c
LOCUS 349980 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 7063 from Patent EP1108790.
ACCESSION AX127147 AX114121
VERSION AX127147.1 GI:14041135.
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 349980)
Nakagawa, S., Mizoguchi, H., Ando, S., Hayaishi, M., Ochiai, K.,
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.
Novel polynucleotides
Patent: EP 1108790-A 7063 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)

```

```

FEATURES
source
1..349980
/organism="Corynebacterium glutamicum"
/db xref="taxon:1718"
/note="Seq 1 to long (3,309,400) split in 11, seq 7063
1..200,001 1..549,980"
BASE COUNT 80727 a 91049 c 97618 g 80586 t

```

## ORIGIN

```

Query Match 100.0%; Score 873; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 4e-185;
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  GTCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGCGCAATCGTGTGACCTGGTTG 60
Db 128548  GTCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGCGCAATCGTGTGACCTGGTTG 128489

QY 61  AACCCGAATGCGTATTTGGACGCGTTTGTGTTATCGCGGCGGTTCGCGCGCAATACGCG 120
Db 128488  AACCCGAATGCGTATTTGGACGCGTTTGTGTTATCGCGGCGGTTCGCGCGCAATACGCG 128429

QY 121  GACACCGGACGGTGGATTTTCGCCGCTGGCGGCTTCGCGGCAAGCTGATCTGGTTCCCG 180
Db 128428  GACACCGGACGGTGGATTTTCGCCGCTGGCGGCTTCGCGGCAAGCTGATCTGGTTCCCG 128369

QY 181  CTGGTGGGTTTTCGCCGCGCAGCAGCATTTGTCAAGCCCGCTGTCCAGCCCCCAAGGTGTGGCG 240
Db 128368  CTGGTGGGTTTTCGCCGCGCAGCAGCATTTGTCAAGCCCGCTGTCCAGCCCCCAAGGTGTGGCG 128309

QY 241  TGGATCAAACTGCTGCTGGCAGTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG 300
Db 128308  TGGATCAAACTGCTGCTGGCAGTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG 128249

QY 301  GGTAGTTTTTCGCCGCTTTTGGAAATCGGTGGCTTCGCCCAAAATGTTGATCGCGGCTCG 360
Db 128248  GGTAGTTTTTCGCCGCTTTTGGAAATCGGTGGCTTCGCCCAAAATGTTGATCGCGGCTCG 128189

QY 361  TGGGAATCTCATGATCGCTCCAACTCGCGGTTCAGAAATCTCCAAGTTGTTGAGTGAA 420
Db 128188  TGGGAATCTCATGATCGCTCCAACTCGCGGTTCAGAAATCTCCAAGTTGTTGAGTGAA 128129

QY 421  TCAAGCTGTTGTTCAGCTGTCTCAACTGACGAAGCAACCAATCAATGATGCTGTCAGGTA 480
Db 128128  TCAAGCTGTTGTTCAGCTGTCTCAACTGACGAAGCAACCAATCAATGATGCTGTCAGGTA 128069

QY 481  TCCGCGCGCTACTCTCTTGTCTCGCGCAGCACCCTGCAAGCGCCATCTGCGCAAGTGAC 540
Db 128068  TCCGCGCGCTACTCTCTTGTCTCGCGCAGCACCCTGCAAGCGCCATCTGCGCAAGTGAC 128009

QY 541  TGCCCGCTTCTTGGCGGATGTCATTGAGCTTTCGCGACCATATCAATATGTTTCAGGTTTC 600
Db 128008  TGCCCGCTTCTTGGCGGATGTCATTGAGCTTTCGCGACCATATCAATATGTTTCAGGTTTC 127949

QY 601  AACATGCCCTCAGACAGGACTTACCTGGCTGGCGGCGGAAACCTCTTGGAATTCATCG 660
Db 127948  AACATGCCCTCAGACAGGACTTACCTGGCTGGCGGCGGAAACCTCTTGGAATTCATCG 127889

QY 661  AGATATTTCTCGTGGCAGCGCCCTGCGCAAGTGTGAGAAAGCAATGACGCCAAGACCA 720
Db 127888  AGATATTTCTCGTGGCAGCGCCCTGCGCAAGTGTGAGAAAGCAATGACGCCAAGACCA 127829

QY 721  TTGTTGGCAGCTGACTGCAACAAAGTTCTCACCGTCATTCGCCCGGTTCTCTCCACCAACGA 780
Db 127828  TTGTTGGCAGCTGACTGCAACAAAGTTCTCACCGTCATTCGCCCGGTTCTCTCCACCAACGA 127769

QY 781  TTAATGATGGAATAGCTTGGCTGATGATCAGAAAGCGGCGAGCCCTCTTCCGCCATGAC 840
Db 127768  TTAATGATGGAATAGCTTGGCTGATGATCAGAAAGCGGCGAGCCCTCTTCCGCCATGAC 127709

QY 841  TCAGCGGCTCCGCTGTGAGCTCTGACCGTAG 873
Db 127708  TCAGCGGCTCCGCTGTGAGCTCTGACCGTAG 127676

```

```

RESULT 5
AX123538/c
LOCUS 627 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3454 from Patent EP1108790.
ACCESSION AX123538
VERSION AX123538.1 GI:14041026

```

KEYWORDS  
SOURCE  
ORGANISM

Corynebacterium glutamicum.  
Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE  
1 (bases 1 to 627)  
AUTHORS  
Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
TITLE  
Novel polynucleotides  
JOURNAL  
Patent: EP 1108790-A 3454 20-JUN-2001;  
KYOWA HAKKO KOGYO CO., LTD. (JP)  
FEATURES  
source  
1..627  
Location/Qualifiers  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
BASE COUNT  
139 a 159 c 185 g 144 t  
ORIGIN

Query Match 65.1%; Score 568; DB 6; Length 627;  
Best Local Similarity 100.0%; Pred. No. 8.9e-117;  
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 306 GTTTTCGGGGTTTGAATCGGTGGCTTCGCCCAATGTTGATCGCGGCTCGTGGGA 365  
DB 627 GTTTTCGGGGTTTGAATCGGTGGCTTCGCCCAATGTTGATCGCGGCTCGTGGGA 568  
QY 366 AATCTATCGATCGCTCCCACTCGCGTTCAGAAAATCCAAAGTTGTTGATGAATCAAG 425  
DB 567 AATCTATCGATCGCTCCCACTCGCGTTCAGAAAATCCAAAGTTGTTGATGAATCAAG 508  
QY 426 GCTGTTTCAGCTGCTCACTGAGGAGGACCAATCAATGCACTGCTGATCCGC 485  
DB 507 GCTGTTTCAGCTGCTCACTGAGGAGGACCAATCAATGCACTGCTGATCCGC 448  
QY 486 GCCGTACTCTCTTCTCGCGCAGCACCATGCAAGCGCCATCTGCGCAAGTGAATGCCC 545  
DB 447 GCCGTACTCTCTTCTCGCGCAGCACCATGCAAGCGCCATCTGCGCAAGTGAATGCCC 388  
QY 546 GCCTTCTCGCGGAGTCAATGAGCTTTCGCGGACCATATCAATATTTGTTCAAGTTCAACAT 605  
DB 387 GCCTTCTCGCGGAGTCAATGAGCTTTCGCGGACCATATCAATATTTGTTCAAGTTCAACAT 328  
QY 606 GCCCTCAGACAGGACTTACCTGCTGGCGGGAGACCTCTGGAATTCATCGAGATA 665  
DB 327 GCCCTCAGACAGGACTTACCTGCTGGCGGGAGACCTCTGGAATTCATCGAGATA 268  
QY 666 TTTGTCGTGAGCAGGCTTCGCGAAGTGGTGAGAAAGCAATGACGCCAAGCATTGTT 725  
DB 267 TTTGTCGTGAGCAGGCTTCGCGAAGTGGTGAGAAAGCAATGACGCCAAGCATTGTT 208  
QY 726 GGCAGCTGACTGCAACAGTTTCTACCGTCAATCCCGGTTTCTCCACCCCAACGATTAAAT 785  
DB 207 GGCAGCTGACTGCAACAGTTTCTACCGTCAATCCCGGTTTCTCCACCCCAACGATTAAAT 148  
QY 786 GATGAATAGCTTGGCTGATCAATCAGAGCGGCGAGCCCTCTCCGCGCATGAACTCAGC 845  
DB 147 GATGAATAGCTTGGCTGATCAATCAGAGCGGCGAGCCCTCTCCGCGCATGAACTCAGC 88  
QY 846 CGCCTCCGCTGTGAGCTCTGGACCGTAG 873  
DB 87 CGCCTCCGCTGTGAGCTCTGGACCGTAG 60

RESULT 6  
AX067087/c  
LOCUS  
AX067087 993 bp DNA linear PAT 24-JAN-2001  
DEFINITION  
Sequence 669 from Patent WO010805.  
ACCESSION  
AX067087  
VERSION  
AX067087.1 GI:12544795  
KEYWORDS  
Corynebacterium glutamicum.  
SOURCE  
Corynebacterium glutamicum

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE  
1 (bases 1 to 993)  
AUTHORS  
Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.  
TITLE  
Corynebacterium glutamicum genes encoding proteins involved in  
membrane synthesis and membrane transport  
JOURNAL  
Patent: WO 010805-A 669 04-JAN-2001;  
BASF AKTIENGESSELLSCHAFT (DE)  
FEATURES  
Location/Qualifiers  
1..993  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
CDS  
101..973  
/note="RXN03164"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC26771.1"  
/db\_xref="GI:12544796"  
/translation="MIYRRVNSGLKLPALISLGLWHNFGDDKPLSTQSIHRAFDGR  
VTHFDLANNYGPAGSAETNFRILREDLKRDELIISSKAGMDWMPYGFSGSRK  
YLVSSLDQSLTRLGLDYDIIFYHRRPDPDPLEETMYALRDIYVAGKALYVSIYGP  
ELTAEAEFMAEEGCPILLIHOPYSYIINRWVEEPDGDGLENLIQSAANNGLGVIAFSP  
AQGLLTDKYLDIPEGSRASQKSLSEGMLNVNNIDMVRKLNIDIAOERQSLAOWALA  
WVLRQREYVAGLP"

BASE COUNT 222 a 247 c 277 g 247 t  
ORIGIN  
Query Match 46.3%; Score 404.2; DB 6; Length 993;  
Best Local Similarity 98.8%; Pred. No. 4e-80;  
Matches 418; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 453 AGCACCAATCAATGCACTGCTCAAGGTA--TCGGCGCGTACTCTCTTGTGCTCGCGCAGC 510  
DB 993 AGCACCAATCAATGCACTGCTCAAGGTAATCCGGCGCGTACTCTCTTGTGCTCGCGCAGC 934  
QY 511 ACCATGCAAGCGCCATCTGCGCAAGTGAATGCGCGGTTCTGCGCGATGTCATTGAGC 570  
DB 933 ACCATGCAAGCGCCATCTGCGCAAGTGAATGCGCGGTTCTGCGCGATGTCATTGAGC 874  
QY 571 TTGCGGACCATATCAATATTTGTTCAAGTTCAACATGCCCTCAGACAGGCACTTACCCCTGG 630  
DB 873 TTGCGGACCATATCAATATTTGTTCAAGTTCAACATGCCCTCAGACAGGCACTTACCCCTGG 814  
QY 631 CTGGCGCGGGAACCCCTCTGGAATTCATCGAGATATTTGTCGTGAGCAGGCCCTCGCGCA 690  
DB 813 CTGGCGCGGGAACCCCTCTGGAATTCATCGAGATATTTGTCGTGAGCAGGCCCTCGCGCA 754  
QY 691 AGTGTGTGAGAAAGCAATGACGCCAAGCATTGTTGGCAGCTGACTGCAACAGTTCTCA 750  
DB 753 AGTGTGTGAGAAAGCAATGACGCCAAGCATTGTTGGCAGCTGACTGCAACAGTTCTCA 694  
QY 751 CCCTCATCGCCCGGTTCTCCACCCCAACGATTAATGATGAATAGCTTGCTGATGAATC 810  
DB 693 CCCTCATCGCCCGGTTCTCCACCCCAACGATTAATGATGAATAGCTTGCTGATGAATC 634  
QY 811 AGAAGCGGCGAGCCCTCTCCGCGCATGAACATGACGCCCTCCGCTGAGCTCTGAGCCG 870  
DB 633 AGAAGCGGCGAGCCCTCTCCGCGCATGAACATGACGCCCTCCGCTGAGCTCTGAGCCG 574  
QY 871 TAG 873  
DB 573 TAG 571

RESULT 7  
AX063771  
LOCUS  
AX063771 822 bp DNA linear PAT 24-JAN-2001  
DEFINITION  
Sequence 53 from Patent WO0100843.  
ACCESSION  
AX063771  
VERSION  
AX063771.1 GI:12541483  
KEYWORDS  
Corynebacterium glutamicum.  
SOURCE  
Corynebacterium glutamicum



```

proteins
JOURNAL Patent: WO 0166573-A 51 13-SEP-2001;
FEATURES BASF AKTIENGESSELLSCHAFT (DE)
source Location/Qualifiers
1..822
/organism="Corynebacterium glutamicum"
CDS /db_xref="taxon:1718"
101..802
/notes="RXA01394"
/codon_start=1
/transl_table=11
/protein_id="CAC88508.1"
/db_xref="GI:15859124"
/translations="MEF1TGLLGLGALLSLGSPQNVLIKQIKREGILAVLLVCL
SDVPLFTAGLVGLDLSNAAPIVLDIRWGGIAYLLWFAVMAAKDAMTKVRAPOIIE
ETPEVDPDTPGLGSAYAVDTRNRVREYVSVDKQVWVZPMLMAIVLTLNPNAYIDA
FVFIQGVGAQYDGTGRWTFPAAGFAAASLIWFLVGFGAALSRPLSSPKVWRWVNVV
AVVNTALATIKMLMG"
BASE COUNT 167 a 192 c 246 g 217 t
ORIGIN

Query Match 37.3%; Score 326; DB 6; Length 822;
Best Local Similarity 100.0%; Pred.No.1.3e-62;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAAGCAGCGGGTTTCGGTAAGCCCATGTTGATGGCAATCGTGCTGACCTGTTG 60
DB 497 GTCGATAAGCAGCGGGTTTCGGTAAGCCCATGTTGATGGCAATCGTGCTGACCTGTTG 556

QY 61 AACCCGAATCGTATTGGACGGCTTTGTGTTTATCGCGCGCGTTCGCGCGCAATACGGC 120
DB 557 AACCCGAATCGTATTGGACGGCTTTGTGTTTATCGCGCGCGTTCGCGCGCAATACGGC 616

QY 121 GACACCGGACGGTGGATTTTCGGCTGGCGCGTTCGCGCGCAATCGTATCGTTCCTCG 180
DB 617 GACACCGGACGGTGGATTTTCGGCTGGCGCGTTCGCGCGCAATCGTATCGTTCCTCG 676

QY 181 CTGTGGTGTTTCGGCGCAGCAGCATTTGTCACGCGCGCTGTCAGCGCCCAAGGTGTGGCGC 240
DB 677 CTGTGGTGTTTCGGCGCAGCAGCATTTGTCACGCGCGCTGTCAGCGCCCAAGGTGTGGCGC 736

QY 241 TGGATCAACGTCGTCGTGGCAGTTGTGATCAGCGCATTTGCCATCAAACTGATGTTGATG 300
DB 737 TGGATCAACGTCGTCGTGGCAGTTGTGATCAGCGCATTTGCCATCAAACTGATGTTGATG 796

QY 301 GGTTAGTTTTCGGCGGTTTGGGAATC 326
DB 797 GGTTAGTTTTCGGCGGTTTGGGAATC 822

RESULT 9
AX123539
LOCUS AX123539 708 bp DNA linear PAT 11-MAY-2001
DEFINITION Sequence 3455 from Patent EP1108790.
ACCESSION AX123539
VERSION AX123539.1 GI:14041027
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
1 (bases 1 to 708)
REFERENCE
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3455 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source Location/Qualifiers
1..708
/organism="Corynebacterium glutamicum"
BASE COUNT 134 a 173 c 221 g 180 t

```





COMMENT FEATURES source	URL: <a href="http://www.kazusa.or.jp/rhizobase/">http://www.kazusa.or.jp/rhizobase/</a> , Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934 On May 11, 2001 this sequence version replaced gi:11994976.	
	Location/Qualifiers	
	1..345783 /organism="Mesorhizobium loti" /strain="MAFF303099" /db_xref="taxon:381" complement(6..884) /gene="mlr3011" complement(6..884) /gene="mlr3011" /codon_start=1 /transl_table=11 /product="transcription regulator" /protein_id="BAB50000.1" /db_xref="GI:14023394"	
gene	/translations="MURKMANSTFIRVIETGSPSAAARHURVQSVASKAVAQLEBRL GVRLSKRTGLSPTDAGLSFYEGAKRANDEADENAAARGTAAALTGLLRIGTSVTF GRLLVPRKAFIDAHPLRDLVEVAMDNDVLDVEEGVDVNLRLGESFAPAHIGRKIAA CRLVVAAPSYLAREGEPIAPDELRSHEALFVRAGTSIRWTFRRGAEEVSLAGQGRLL RVTSAGIRREAVLSGIGFTVGSSEMFDAELRSRAVRPILTEWTLPPVELWALPPAGRK ASARATFEVFLAGSL" 994..1740 /gene="mlr3013" 994..1740 /gene="mlr3013" /codon_start=1 /transl_table=11 /product="short chain dehydrogenase or 3-ketoacyl-CoA reductase" /protein_id="BAB50001.1" /db_xref="GI:14023395"	
	/translations="MGKLNKVAIVTGAAGIGAGIAGIAKRLAELAEAGAAVVVNYATSRRAA ERTVGIYAGKATAVQGVDAEAEVEGFARAEIAPGRPSILINNAATASFGTFEE ASADTRRLDFTNVLGITLTKAAMRHPFETGGSIVNIGTISWNPVPNTLSYASKA AIDTLSLSRELGARNIRNIVAPGYTHTEMTGVMGTDGFMNLIAGVPLQGRFGKP DDIAPTAVFLADEAAWLITGERINASGGAR" 1914..3371 /gene="mlr3014" 1914..3371 /gene="mlr3014" /codon_start=1 /transl_table=11 /product="homoserimidine synthase" /protein_id="BAB50002.1" /db_xref="GI:14023396"	
	/translations="MLDKMANENWPVYGITGVPMVIMGFSIGRGTPLIERHFKFDK SRMTLDPRDTRKLLDERGIAFVQBAVTENNYKKLLTLLTNGGGQGFVNLSDVTG SVDMRLCKRLGLYIYDVTVPWLGIFYFOAKADNASRTNVALREAMIKEKHDPGGAT VSTCGANPMVSWFVKQALVNLATDLGLEFSEPAOEDREGWAKMKKAGVKGITAE RDQRTKKPKPMDVFNWTSVEGFISEGLOPAELGWTHKMPKNGKKGKHSKRAI YLEPGANTVRSCWCTPGCAQGLLVTHNEAISADPFTVRSKGKVQYRPTCHYAIH PSNDAMLSDEMFGAAGKQPVHHVLDENELVDGDELGLVLYGHDKNAYWYGSQSLSL AEARKLQYONATGMQVTSAGVLAGVMVALENPDAGTVEADMDYKRCLDVQSPYILGPV KGYTDTWTLDRPPLGFLPEDLDRSDPWQFRNILVR" 3431..3931 /gene="mlr3016" 3431..3931 /gene="mlr3016" /codon_start=1 /transl_table=11 /product="outer membrane lipoprotein" /protein_id="BAB50003.1" /db_xref="GI:14023397"	
gene	/translations="MRPGAACVLRDARPFQHFGRSHRAQCNRRFLDRCVWRGFQWTI ARKVLSIGMAGSFAMMLAACTTSGDPAPPMAAAPKGVESWIDAKGTGLSTFTAGKFA TVATDTGQXLDAGSYTMTGATSVETNGTSLIRQT PVSFNCLLISTSQTSSGQNF VLTTRRT" 4123..6243 /gene="mlr3017" 4123..6243 /gene="mlr3017"	
	/EC_number="3.1.1.3.5" /codon_start=1 /transl_table=11 /product="5'-nucleotidase" /protein_id="BAB50004.1" /db_xref="GI:14023398" /translations="MKKLAIAALSVTLGLSAGASFADYTLNLIHFNDWHSRIKGN KYESTCSADEETKGCIGAGRLITAJAQRKELSGONVLLNAGDSFGQSFLPYTYTK GTVEEFNLQMPDANTLGNHEPDDGESALYLDKAKFPVIVSANWPNDRSGAAGKI KFSIVVEGGQKIGVAVNTDPELASPGNIAADDESITAEVEKLKAGGVANKII AVTHIGYNRERDIIAKIPGVVVVGGSHSTLLSNTDPRKAAGPYPTMVDNPGYKVPVW QAASYSKYLGEFKVFDNDGNGVKSADPILDKSITDPPAVLARIKELGASIEALKKN KEVAETDVIDGSRNCRACECMGNLVSDAILDRVKQGVVEIVISNGGLASIDKG VTVMGEVLVLPFONLATFOISGKDLVAGLESLSQIEDGAGRPOVAGLKYSPDKS VAPNAGRVKSVEMWENGAWTPIKPKDYLIVATNVTYRGGDGYKYFAERAKNAYDGP GLEQVADYLGHRPITFPLDGRITETIAATVAAPAEPAKPAEATPAPAPAPAKPA EAAAPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPAEPAKPA IYEANKGQRPRLTTPGATLTIPAK" 6381..7442 /gene="mlr3019" 6381..7442 /gene="mlr3019" /EC_number="4.99.1.1" /codon_start=1 /transl_table=11 /product="ferrochelatase" /protein_id="BAB50005.1" /db_xref="GI:14023399"	
	/translations="MTSAIADVDPKAAPVGTAGHAPAGAGKGVLLVNLGTDPDGTEF KPMRYLREFLSDPRVIELNKAIWYPIYGLVLTTRPKKSGANVARIWNERNESPLR TYTRAQSEKLAELGDLDPVTVDMWARYGNPSTASVAERLVAQGCDRILSPFLYPOYS ATTAYANDQFLRALMKLRAPATNSVPPIYAEVPIEALASSIERHLATLDPEFVV ITSYHGIPKPYSDKGPYQAHCELTLLRLREKLGWDEKGLITTFQSRFGAGLEWLPYTS DVKYELKGGKVSIAIYVNGFVSVDCTELDEIGREAAETFFHAGKKNFAHITPCLNDS REGMTVIAMVRRRELSGWV" 7615..8826 /gene="mlr3020" 7615..8826 /gene="mlr3020" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB50006.1" /db_xref="GI:14023400"	
gene	/translations="MPRKLDLRTGRPWSAYRAPAVPTSALTTRDVTVDLIIIGMGISG AMAEALFADGSHVICIDRRGLKXSTAATTAATVQFEIDQDPLSTLSKRMIGKAAQAAW RRSRIASNLAGRIEDLAIDCRLARTPSLYLAGTLLGPAEMRDEAEARROAGVAATYL TPQVLAQTFGIDRGGAIIISHGNIALDPRKLTAGLLKALKERKARLYAPAEATAIEDSA GEAVVATKDGPTITARHVLTATVELVDIVPAAAHRIISTWAIATCPOPRKLNWPGPAP IWEASDPYILURATADGVIICGGDEDDPADSTRDELADKSARIADKGLRFLFPYLDV RPEFAWTSQFSTAGLPIYICAVPGHPRIHAVMGYGGNGITTSRIASIEIISASIKGSA DTDALFAPNR" 8948..9898 /gene="mlr3021" 8948..9898 /gene="mlr3021" /note="hypothetical protein" /codon_start=1 /transl_table=11 /protein_id="BAB50007.1" /db_xref="GI:14023401"	
	/translations="MDFSGFDIAIGALLVLVLLIKGIRTIPOQYNTVTRFGRYTK TSLPGLNPIFFPVDRIKAKNMMEQVLDVPSQEIITRDNAIVGVDTGAFFQITLAAQA AVYSGGLQNPAILNLTMTNITVMGMDLDELNSRDINERLIVDERVDEWPGIKIT RVEYKQNDPANIETSMGRWTAEKNKRAQILAAEGLKQSOILAEABGRKEAFAFDAAE RERSAEARATQVVSRAISKGVDQALNYFAQKYTBALGKIGTATNSKIVLMFPEAS SLISLGGIGGIAKEVFRSEGTTCQAQQAARPPVVRSEN" 9912..10376 /gene="mlr3022" 9912..10376 /gene="mlr3022" /note="hypothetical protein"	

```
/codon_start=1
/transl_table=11
/protein_id="BAB50008.1"
/db_xref="GI:14023402"
/translations="MFDRITSELSGPNWMLGVLLVMBIIAPGIFMLWIGIAALIIG
AVSLLIWDAITWQVLAFLSLVSAVVGKLMGGSDQDTPQLNRRGAQWVR
MATLAPIKDGGRGRIKLGDTLWRVSGPDLPACTQVVRTSAUTDLELTVRV"
complement(10629..11630)
/gene="ml13023"
complement(10629..11630)
/gene="ml13023"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAB50009.1"
/db_xref="GI:14023403"
/translations="MHARSLDKKPLDRQASIDSALRTVATEQAGIAALAELENGLAA
PFAQAVDMISKIRGLIVTCVGKSGHIGSKIAATLASTCTPAFFVHPVEANHGDLMGI
AKDDAIIAISWSESKEMLGIVAYSRPFIPIANTVSGTSSALARAADVLLLPRTPE
ACPHGLAPTSTLLQVLIGDALAIALLARGFTPDHFRTHFGQGANLTVWSEIMR
VGQMLPLAVLTQKQPEAVMTLSQKVGCVLIVDANGELAGIITDGDVLAHLRNLAADV
IVDEVMTPTKVPDTPOTLAGTATALLNEHNIGALVVRNNMPLGVVHFHLLRIGAA"
11830..13626
/gene="mlr3024"
11830..13626
/gene="mlr3024"
/notes="unknown protein"
/codon_start=1

Query Match      16.23;  Score 141.2;  DB 1;  Length 345783;
Best Local Similarity 57.08;  Pred. No. 2.5e-21;
Matches 319;  Conservative 0;  Mismatches 208;  Indels 33;  Gaps 2;

QY  314  GGGTTTGGAAATCGGTGGCCCTTCGCCAAATGTTTGATCGCGCGTCTGGGAAATCTCAT 373
      |||||
Db  249376  GGCCTTTGGCTCGCGGAGGCTGCCACAGATGATGTGCGATTGCGGTGGTAGCTAT 249435

QY  374  CGATCGCTCAACTCGCGCTCAGAAACTCAAGTTGTTGAGTGAATCAAGCGTGTGT 433
      |||||
Db  249436  CGATCTCGGCNAATTCGGCATCGCTGAAGTCGAGCACCTTTGAGGCGCGCCGACGAGTCT 249495

QY  434  CCAGCTGCTCAACTGACGAGCACCAATCAATGACCTGCTCAGGTCATCGCGCGTACT 493
      |||||
Db  249496  CGACTGCTCGCGCGCTGCGCGCGATCGAGCGCGAGCTCACCTTACC----- 249544

QY  494  CTCCTTGTCTCGCGCAGCACCCATGACGCGCCATCTCGCGAAGTGACTGCCCGCGTTCCT 553
      |||||
Db  249545  -----CTTGGCGCAGCACCCAGCGCGGCATCTCGCGCAGCGTCTGCGCGCGCTTCG 249597

QY  554  GGGCGATGTCTAGCTTGGGACCATATCAATATTGTTCAAGTTCAAGTCCCTCAG 613
      |||||
Db  249598  CGGCGATGGCGTTTGGAGCGCCCTGATGTTGGCGATGTTGCGGTCTGATGAAGCGCGACT 249657

QY  614  ACAGGACTTACCTGGCTGCGCGGGAACCTCTGGAATTCATCGAGATATTGTCCG 673
      |||||
Db  249658  TCACGACTTGCCCTGGAGGACAGCGTGCCTTCGGGATGCGCCCAAGATATTGTTCGG 249717

QY  674  TGACGAGCGCTCGCGAAGTGGTGAGAAAGCAATGAGCCCAAGACCACTTGTGGCAGCTG 733
      |||||
Db  249718  TCAGCATGCCCTGGGCCAGCGCGGAGAACAGATCGAGCGCACACCGAGCCCTTCCAGCG 249777

QY  734  ACTGCAACAGTTCTCACCGTCATCGCCGCTTCTCTCCACCAACGATTAATGATGAAT 793
      |||||
Db  249778  TGTGAGCAGGCGCGTC-----TTCTCTCCACCCAGCGGTTTCAGCATCGAAT 249822

QY  794  AGCTTGGCTGATGAATCAGAACGCGGAGCGCTCTCTCGCCATGAATCAAGCTCAGCCCTCCG 853
      |||||
Db  249823  AGCTCGCTGTGTGATGACGAGGCGGTGCGAGCTGCGGCAATATGTGCGCAGCCTCGC 249882

QY  854  CTGTGAGCTTGGACCGTAG 873
      |||||
Db  249883  GGGTGGCGCTGCGAATTGTAG 249902
```

RESULT 12  
AE008076/c  
LOCUS  
DEFINITION  
Agrobacterium tumefaciens str. C58 circular chromosome, section 134 of 254 of the complete sequence.  
ACCESSION  
AE008076 AE007869  
VERSION  
AE008076.1 GI:15156608  
SOURCE  
Agrobacterium tumefaciens str. C58 (Cereon).  
ORGANISM  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.  
REFERENCE  
1 (bases 1 to 8472)  
Hinkle,G., Slater,S.C. and Goodner,B.  
Complete Genome Sequence of Agrobacterium tumefaciens C58 (Rhizobium radiobacter C58), the Causative Agent of Crown Gall Disease in Plants  
JOURNAL  
Unpublished  
REFERENCE  
2 (bases 1 to 8472)  
Hinkle,G., Slater,S.C. and Goodner,B.  
AUTHORS  
Direct Submission  
TITLE  
Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street, Cambridge, MA 02139, USA  
JOURNAL  
FEATURES  
source  
1..8472  
location/Qualifiers  
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"  
/strain="C58"  
/db\_xref="taxon:181661"  
79..1077  
/gene="AGR\_C\_2805"  
79..1077  
/gene="AGR\_C\_2805"  
/note="AE005408) putative adhesin"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2805p"  
/protein\_id="AAK87312.1"  
/db\_xref="GI:15156609"  
/translation="MHFMKSLIPLMASAALAASVAGATAAPDVVVSIKPVHSLVAAL MKGVGEPLIVDGAASPTHYLNPSNARKLEKADVVFWVGPLEAFLOKPLEALASKA TVLEDAKGLKLPFRKGGPPEAHDDCEGHEAHAGTEDEGADHNDGNDHAGSEHE HGAYDTHLWLPANNAQALETALIAADAGNATYQANTKKLIDDLDAEDAEVET VPAKDKPFIHFDAIYQIFEHRYGVKTAGSITVSETLPDGRVKOMGKVRQLGAPC VFAEPQEPKLVSVITEGTAASATLDPAAATLTGPDLYPKLMRGIAGSLKNCUS"  
complement(1099..2166)  
/gene="AGR\_C\_2807"  
complement(1099..2166)  
/gene="AGR\_C\_2807"  
/note="TRANSCRIPTION REGULATOR AGLR"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2807p"  
/protein\_id="AAK87313.1"  
/db\_xref="GI:15156610"  
/translation="MWRNMDMTGKSGGDMPSMGRRPTLKTIAVTGLGTITVTSRA LKADPDIGATKERVRLIARQIQYQPNRAGVRLRTKTNVIALVLSVDEELMGFTSQM VFGITVSLTQYHLVTPPHIAKDSMVPYIRILETSGADGVIIISKIPNDRVRPMT ENMPFVTHGRSDMIGIEHAFDFDNEAYAVEARLAOCGRKRIAVIIPPSPRFSFDH ARKGFNRGIRDPGLTEFPIDAVTIEPLEKIRDFQRLMOSSDRPDGIVISIGSSSTIA LVAGFEAGVKGIDVDIVSKQSAEFLNWKPIQIHVNEIDIKLAGRELAKALLARNG EAAETLQSIQSGPWSMAPKP"  
2640..5138  
/gene="AGR\_C\_2809"  
2640..5138  
/gene="AGR\_C\_2809"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/product="AGR\_C\_2809p"  
/protein\_id="AAK87314.1"  
/db\_xref="GI:15156611"  
/translation="MQGIMISSSTPETVIDLAGLWHLASVEGDHATEISIPGDIHSA

```

LKNAAIIPDPYHGANEKAVQWVAQQDWIERTFILDDAEASWYLDIDYLDYVAIVFN
DVPVLSADNCFRYPEDISRAVPGENTIRIHFNISNITAGAEQAQPPFIPYHPGNS
PIANGNLRFCQFQWDMNIAIAPGLYGLIKLLKRLDITARIEHVVSQSHVGGVGL
HVAUTLPAELGPASLPVYLSGLDERLRELCGVAGETVVRHVPFVNDPDLWPPAGSEQ
TLVLTJVELDETTRQIGPRTIELLTDKDEAGSRAFRINGREIFCRGNWIPADAL
YSLTSREKTELLCSAVEANNNIRVWGGGYEEDFYDLCDRLGLLVQDFACNL
YPCSEDFLDNVEHVDYQVRLSSHSFIALWCGDNLVGLATWFDSSRNNRDRYLVAI
DLNRNTEKAKATPEALWPPSPASGYLDYGDWADHAGDGMWYVSHENKSPDN
YHQVPRFCSEFQSYTSPNVRTVAEDKDMNIAFPVIELHOKYVGGNRIAGTWFR
YFQKRWALHYAARFQFQVAVSALPSADGRVTFSMUNDTAEVDIMNIVALMDGN
RVPLKANGCTSDKKAATLTDIMDSLPDGAILLANFPIASNGMTGGHVRDITYKALE
LQPAGLEFSVGPLUNGQFELDVAAGLALFIMLEADQOPGRYSNDLFDLAAGETRRRIIF
TPKAGGQPHFRIPDLHTCOSSPNPGIETWRRKA"
gene
5278. 6273
CDS
/gene="AGR_C_2811"
/gene="AGR_C_2811"
/notes="hypothetical protein b3001 - Escherichia coli
(strain K-12)"
/codon_start=1
/transl_table=11
/product="AGR_C_2811p"
/protein_id="AAK87315.1"
/db_xref="GI:15156612"
/translations="MKYNHCGKTKLKLPAISLGLMHNFGNDTPTHQTKQAIRAFDLG
ITFDLANNYPGPGSAETAFGLKTLDFRGYRDEMIISKAGYNNMPQPYGEMGRK
YLLSCDQSLKRLMGLDVIDFYSHRFDNPLTEETCGALDQIVRSCKALYVLSYNS
KRTREAAILLKDTGTCILHQPSYMINRWIEDGLVDLIELGIGSIVFSPLAGML
TTYKLGVPDGRASQSKSLNPAFLNERNVENIRALNSIAERRGQTLAQMALAWLRG
GRITSALIGASRVEQEDCVKALDNEAFSELAETDRYAKDADINLWAKSAERV"
6465. 6791
gene
6465. 6791
CDS
/gene="AGR_C_2812"
/gene="AGR_C_2812"
/notes="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_C_2812p"
/protein_id="AAK87316.1"
/db_xref="GI:15156613"
/translations="MKFVTIILAAVTLLAAPWAQASRHHDRHGVTVVERQVTTKKVI
VKHRDRGQRUSARERNRNVDRYRHYLAEPDRDQWRVVDNQFLINAVSGLIV
GLAAAR"
6465. 6791
gene
6465. 6791
CDS
/gene="AGR_C_2814"
/gene="AGR_C_2814"
/notes="6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING"
/codon_start=1
/transl_table=11
/product="AGR_C_2814p"
/protein_id="AAK87317.1"
/db_xref="GI:15156614"
/translations="MEQAEIGLIGLVMSGNLALNIAEKNKIIVFNRTPEVTRKPYA
DAGELQIIPCEITEEFVAARPPRIIMIKAGDPVQDQMEILKPLHANGDIMIDA
GNANFRDTRRPNLKSGLTPIGMVSGEGRGARGPSIMVGGSTEDSKRVEKVLTS
ISAKYNDPDCVWLNGDAGHGVKTHNGIEYADMQMIAEYIGLRDLGKMSAVEIAD
FVARNKGRNLSYLBITEKVTARADPTTKKEMVDLIDKACQGTGKMSVLEAQNMG
VAATAERAARAKSSQDEREAEEKIFGLPTLAAAPADKFAFTADLESALLAAKVG
AYAQGFPAISAKSFNNWNLPMPTIARIWAGCIIRSEFLDITSAFTKDPHVNLIIV
TAFSAIVKDTDAPLRVVSVALSGLPVLSALAGLYPDAYRRGRGSANLIQAQRDF
FGAGFPRTDGVDPKHPGWSGADIP"
BASE COUNT 1814 a 2570 c 2348 g 1740 t
ORIGIN

```

Query Match 15.6%; Score 136.6; DB 1; Length 8472;  
 Best Local Similarity 56.8%; Pred. No. 3e-20;  
 Matches 312; Conservative 0; Mismatches 204; Indels 33; Gaps 2;

```

QY 325 TCGGTGGCCCTCCGCCAAATGTTGATCGCGCGCTCGTGGGAAATCTCATCGATCGCCTCC 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6263 TCGGCAGATTTTCCCGCAGAGGTTGATATCCGCATCTCTTGGCGTAAACGGTTCGATTCGGCC 6204

```

```

QY 385 AACTCGCGCGTCAGAAAATCCAAAGTTGTTGAGTGAATCAAGGCTGTTGTCAGCTGCTCA 444
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6203 AGTCTCTCGGTAGAGAACTCGGCATTTATCGATGCTGTTTCAGCAGACCTCGACCTGTCG 6144
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 445 ACTGAGAAAGCACCACCAATCAATCGTGTGTCACGGTATCCGCGCGCTACTCTCTCTGCTCG 504
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6143 ACACGGCTTGGCGCAATCAATGCTGAGGTAATGCGGCC-----GCCG 6102
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 505 CGCAGACCCATGCAAGCGCCATCTCGGCAAGTACTGCGCGGTTCTCTGGGGGATGTCA 564
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6101 CGCAGAACCCAGGCAATTCGCATCTCGGCCAGCGCTTGGCCACCGCGCTCGGCAATGCTG 6042
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 565 TTGAGCTTGGCGACCATATCAATATTGTTTCACTTCAATCGCTTCAACATGCCCTCAGACAGGACTTA 624
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6041 TTGAGCGCGCGGATGTTTTCGACATTCGCTGTTGAGGAAGCGCGGTTGAGTGACTTG 5982
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 625 CCTGCTGCGCGGGAACCTCTGGAATTCATCGAGATATTTGTCTCGTGAGCAGGCC 684
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5981 CTCTGTGAGGCGACGGCTGCCATCGGCACACCGCCAGATATTTCTGCTGAGCATGCC 5922
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 685 TCGGCAAGTGTGAGAAACAATGATGCGCAAGACCAATGTTTGGCAGCTGACTGCAACAAG 744
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5921 TCGCGCAGCGCGCAAAAGACGATGGAGCGGATACCCAGTTCTTCCAGCGTAT----- 5870
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 745 TTCTCACCTGTCATCGCCCGGTTTCTCCACCAACGATTAATGATGAATAGCTTGGCTCA 804
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5869 -----CGACAGACCGCTTCTTCGATCCAGCGGTTGATCATGAATAGCTCGGCTCG 5817
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 805 TGAATCAGAAAGCGGCGAGCCCTCTCCGCGCATGAATCAGCGCGCTCGGCTGTGAGCTCT 864
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5816 TGGATGATCGAGGCGTGGCGAGATCTTTCAGGATAGCGCGGCTCGCGGTGCGCTTC 5757
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 865 GACCGGTAG 873
|||||
Db 5756 GAGTTGTAG 5748
|||||

```

RESULT 13  
 A8009111/c

LOCUS  
 DEFINITION

AGrobacterium tumefaciens strain C58 linear BCT 20-DEC-2001

137 of 256 of the complete sequence.

ACCESSION AE009111 AE008688

VERSION AE009111.1 GI:17739946

SOURCE Agrobacterium tumefaciens str. C58 (U. Washington).

ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

REFERENCE 1 (bases 1 to 13051)

AUTHORS Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,

Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,

Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,

Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,

McLellan, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C.,

Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P.,

Zhang, S., Yoo, H., Biddle, P., Jung, M., Krespan, W., Perry, M.,

Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M.,

Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.

The genome of the natural genetic engineer Agrobacterium

tumefaciens C58

Science 294 (5550), 2317-2323 (2001)

21608550

11743193

REFERENCE 2 (bases 1 to 13051)

AUTHORS Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,

Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F.,

Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G.,

Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M.,

McLellan, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C.,

Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I.,

Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.

# Direct Submission

Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

## FEATURES

### source

### Location/Qualifiers

1. .13051

/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"

/strain="C58"

/db\_xref="taxon:180835"

83. .1072

/genes="znuA"

/note="Atul521"

83. .1072

/genes="znuA"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=1

/product="ABC transporter, substrate binding protein [zinc]"

/protein\_id="AAL42526.1"

/db\_xref="GI:17739947"

/translation="MKSILIPLMASALAASGATAADPVVVSIPKPVHSLVAIMKG

VGEPLIVGASPHYNLRPSNARKLEKADVFWVPGLEAFLOKPLEALASKATV

ELEDAGKELPRKGGPEANDDGEAEHAGHTEDEGADHNDHAGSEHEHGA

YTHLMDPANAKAQALETALIDAGNAATYQANTKKGIDDLALDAVEVETVKP

VKDPFIVPHDAYQYFERYGVKTAGSITVSPETLPQADRVKQMEKVRQLGAPCVFA

EPQFEPKLVITEGTAAKSATIDPEATLTGPDLYFKLMRGIAGSLKNCLIS"

complement (1094. .2113)

/genes="Atul522"

complement (1094. .2113)

/genes="Atul522"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=1

/product="transcriptional regulator, LacI family"

/protein\_id="AAL42527.1"

/db\_xref="GI:17739948"

/translation="MFSAMGGRTLTATYMTGLTITVSRLKADPDIGAEKERV

LIARQICYPNAGVRLRTKNTVIALVLSVDELMGFTSQMVGITVLSITQYHLV

VTPIHAKDSMPVIRVILETSADGVIISKIBNDPRVFRMTERNMPFVTHGRSDMI

EHAHPDNEAYEAVEARLAOCGRKRIAVIPSPFSFDHARKGFNRGIRGLTE

FPIDAVTIEPLEKIRDFQRLMQSSDRPDGIVISGSSSTIALVAGFEAGVYKIGEDV

DIVSKQSAEFLNWKIPQIHTVNEDIKLKGLRELAKALLARINGEAETLSIGPVWVS

MAPKP"

2650. .5133

/genes="mana"

/note="Atul523"

2650. .5133

/genes="mana"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=1

/product="beta-mannosidase precursor"

/protein\_id="AAL42528.1"

/db\_xref="GI:17739949"

/translation="MISSSTPETVIDLAGLHLASVEGDHATEISIPGDIHSLKNA

IIIPDVHANEKAVQWVAQODWIERTFILDDAASWYLDIDYLDVAVFYVNDVPL

SADNCPREYDTSRAVPCGNTIERTHFSNITAGAEQARQPPYIPYHPGNSPIANG

NMLRKQCHGNDWNTAIPGLYKILLKRLDARIEHVVSQHHVGGVELHVAVT

LFAGPASPVLUSLGDRLURLECGVAGETVVRHVFFENPDLMWPPAGSGEQLYKL

TVELPDVETVQIGFTIELLTIDKDEAGSFAPFRINGREIFRCGANNIPADALYSLTS

REKTEDLLCSAVANMMIRVWGGEYEDFYDLCDRLGLLVQDFMFCNLYPCSE

DFLDNDVHEYDQVRLSHPSIALWCQDNELGALTWFDESRNNDRLVLYADRLNR

TIERKALKATPEALWPPSSPASGLVDGDAWHADGSGDMHYSVSWHENSFDNYHVK

PRCSFEGFSYTSMPVIRTYAEDKDMNIASPIELHQRKNVGNRIAGTMFYFRFP

### gene

### CDS

/gene="Atul524"

5237. .6268

/gene="Atul524"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAL42529.1"

/db\_xref="GI:17739950"

/translation="MVWQPAENRYASMKYNHCGTKGLKLPALISGLWHNFGNDTPHOT

KOATCRRAPDLGITHFDLANNYGPFGSAETAFCGLIKLTDPRGYRDEMISSKAGYNN

WPGYGEWSGRKYLISSCDQSLKMGLDYVDIFYSHRPDPNTPLEETGALDQIVRS

KALVGISSYNSKRTREAAALKDLGTPCIIHQPSYSMINRWIEEDGLVDLIEHLGIG

SIVSPLAQGLMTTKYLGVPDGSRSQSLNPAFLNERNVENRINALNSIAERRGQT

LAQMAIAWLVLRGRITSALICGARVEQVDCVKALDNABFSTELAEIDRYAKADAIN

LWAKSAERV"

6460. .6786

/gene="Atul525"

6460. .6786

/gene="Atul525"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="conserved hypothetical protein"

/protein\_id="AAL42530.1"

/db\_xref="GI:17739951"

/translation="WKKEFTVTLAATVLAAPMAQASRHHDRHGVTVVEROVTTKKVI

VKKHRWDRGQRLSARERNMVDYRRYHLAEPRDQVRVVDNQFLINAVSLIV

GLAAR"

complement (6853. .8283)

/gene="gnd"

/note="Atul526"

complement (6853. .8283)

/gene="gnd"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="phosphogluconate dehydrogenase, decarboxylating"

/protein\_id="AAL42531.1"

/db\_xref="GI:17739952"

/translation="MEQAEIGLIGLVGMSNLALNIAEKGNKIAVFNTPVTRKFYA

DAGELQGIIPCTIEEFVAIRPRPIIIMKAGDPVDQMQMEILKPLHGLNDIMIDA

GNANFRDTRRFDNLKDSGLTFIGWVGSGEGARHGFSINVGGTDSWKRVKVLTS

ISAKYNDPCVAVLNCNDGAFHVKTIHNGIEYADQMIAEYIGILRDLGKMSAVBIAD

VFAENWKGRLNSYLIEITEKVLRAADPTTGKPMVDLIDKAGOKGTGKWSVEAONMG

VAATAIEAARVAARILLSSQDERAEKIFGLPTLAAAPADKKAFTADESALLAAKVG

AYAQFALMSAASKEFNWNLPTIARIWRAGCIIRSEFLDEISAFKDPKHVANLIV

TPAFSAIVKTDAPLRVVSXVLSGLPVSALASALGYFDAYRRGRGSANLIQAORDF

FGAHGFERTDGVKPHGPGSGADIF"

complement (8435. .8596)

/gene="fixS"

/note="Atul527"

complement (8435. .8596)

/gene="fixS"

/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer"

/codon\_start=1

/transl\_table=11

/product="nitrogen fixation protein FixS"

/protein\_id="AAL42532.1"

/db\_xref="GI:17739953"

/translation="MNMLIYLPVALFLGALGLFAPLWSVRSQYEDMDGAARVIED

GDDRPRPS"

complement (8593. .10884)

CDS	/gene="fixl"
	/note="Atul528"
	complement(8593..10884)
	/gene="fixl"
	/notes="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
	/codon_start=1
	/transl_table=11
	/product="nitrogen fixation protein FixI"
	/protein_id="AAL42533.1"
	/db_xref="GI:17739954"
	/translation="MSCCAPGTSGSLGEPLNPPSEELLLASRDLCGLRQTDLSLVPGVYCGACLTITIALGRGQVERARVNLSKRVAVWKEEVNGVTDPADIARAILATGYRIHLFSAGDASTALKRSQLIRVALCFGASANIMLISVSVMSGADNRDMFHHSAMIAPALIYGGRFYSANNAKLHRTNMVDPIALAITLSYALSWEITHRGHEHWFDATSELLPFLRLTGDTHIMDKARSATSALGRPVATIDADQVRDYRLADIEFGMSIAIAGDRIVADLVAVCVGSDDLDMISVNGESAPRVVAGDSLQAQTNLUTGSLFARVTKASCSFLSEVLGLMEASGRARYRIADRAASYSPVVHLLALVTFILGWGIFGGDKOAMLIATAVLIIITPCALGLAVPVVVAAGRLEPRGIMKEGSMERLSRIDTVLPDKTGTLTVCKPRIVETADVKPALMAIAGLAASHRHPLSLHAAYANGALPAYEAVREIPSGVEATEAGTYRLGNRRFACPDETRINNDARKSEVVLSDGRHLASFGFQE
	Query Match 15.6%; Score 136.6; DB 1; Length 13051;
	Best Local Similarity 56.8%; Pred. No. 2.9e-20;
	Matches 312; Conservative 0; Mismatches 204; Indels 33; Gaps 2;
Qy	325 TCGETGCCCTTCGCCAAATTTGATCCGGCGTCTGGTGGAAATCTCATGCATCGCCTCC 384 
Db	6258 TCGGCAGATTTCGCCAGAGTTGATATCCGCATCTTCGGCGTAACCGTCGATTCGCCC 6199 
Qy	385 AACTCGCGCTCAGAAACTCCAAGTTCTTGAGTGAATCAAGGCTGTGTCCAGCTGTCTCA 444 
Db	6198 AGCTCTCGTAGAGAATCGGCATATTCAGAGTCTTTTCACGACGTCTCGACCTGTTCG 6139 
Qy	445 ACTGACGAAGCACCAATCAATCGTCACTGTCAGGTATCCGGCCGTACTCTCTTGTCTCG 504 
Db	6138 ACACGGCTTCGGCCATCATGCTAGGTAATGCGGCC-----GCCG 6097 
Qy	505 CGCAGACCATGTAAGCGGCATCTGCGCAAGTACTGCCCGCGTTCCTGGCGGAGTCA 564 
Db	6096 CGCAGAACCCAGGCAATTGCCATCTGCGCCAGCGCTTCGGCACCGCTCGGCAATGCTG 6037 
Qy	565 TTGAGCTTGCGGACCATATCAATATCTTCACGTTCAACATGCCCTCAGACAGGACTTA 624 
Db	6036 TTCAGCCCGCGATGTTTTTGACATTCGGCTCGTTCGAGGAGCGCGGTTCAGTGACTTG 5977 
Qy	625 CCTTGGCTGGCGGGGAACCTCTGGAAATTCATFCGAGATATTGTCCGTGAGCAGGCC 684 
Db	5976 CTCTGTGAGGCACGGCTGCCATCCGGCACACCGCCACAGATATTCGTGTCGACATGCC 5917 
Qy	685 TGCGCAAGTGGTGAAGAAGCAATGACGCCAAGACCATTTGTTGGCAGCTGACTGCAACAAG 744 
Db	5916 TGCGCCACGCGCGAAAAGACGATGGAGCCGATACCCAGTTCTTCCACGCGTAT----- 5865 
Qy	745 TTCTCACCGTCATCGCCCGTCTCTCCACCCACGATTATGATGGATAGCTTGGCTGA 804 
Db	5864 -----CGACAGACCGTCTTCTTCGATCCAGCGGTTGATCATCGAATAGCTCGGCTGG 5812 
Qy	805 TGAATCAGAAACGCGGCAGCCCTCTCTCGCCATGAACCTCAGCCGCTCCGCTGTGAGCTCT 864 
Db	5811 TGGATGATGACGGCGGTGCGAGATCTTTCAGGATACGGCGGCGCTCGCGGTGCGCTTC 5752 
Qy	865 GGACCGTAG 873 
Qy	5751 GAGTTGTAG 5743 

**RESULT 14**

SC7H9/C  
LOCUS SC7H9  
15586 bp

DEFINITION	SC/H9	I5586 bp	Linear	BCT 12-MAY-2002
SCC33			DNA	
Streptomyces coelicolor cosmid 7H9.				

ACCESSION AL450223 AL645882

VERSION AL450223.2 GI:20520842

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

KEYWORDS	SOURCE ORGANISM	REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED	REFERENCE AUTHORS	REFERENCE AUTHORS	JOURNAL TITLE
----------	-----------------	-------------------	-------	------------------------	-------------------	-------------------	---------------

COMMENT

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBRC and Beowulf Genomics. Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. ([URL: http://www.sanger.ac.uk/Projects/S\\_coelicolor/](http://www.sanger.ac.uk/Projects/S_coelicolor/)) CDS are numbered using the following system eg SC7B7.01c. SC (s.coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4769-4778(1994)) and the Framemaplot program of Bibb et al., (Gene 30:157-66(1984)) as implemented at

<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 7H9 lies between and overlaps with cosmids 5F1 and 2H2 on the AseI-A genomic restriction fragment.

**FEATURES**  
**SOURCE**

```

1. 15586
   /organism="Streptomyces coelicolor A3(2)"
   /strain="A3(2)"
   /db_xref="taxon:100226"
   /clone="coemid 749"

```

misc feature

```

-
stem_loop
/note="nominal overlap with Streptomyces coelicolor cosmid
St7H9
nominal overlap with Streptomyces coelicolor cosmid St5F1"
complement(583..618)
/note="possible stem loop. Score 51: 17/17 (100%) matches,
0 gaps"

```









```

predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE EXOEXOXYRIBONUCLEASE III PROTEIN"
/protein_id="CAC46228.1"
/db_xref="GI:15074583"
/db_xref="SPTREMBL:Q92PT1"
/translations="MKIATYVNGVNGRLGVLRLWLEASPDVVCLOELKAPDPKPPV
KATEAGYGAIMGQKSWNGVAIARDREPTLRKGLPGDDPDTHSRYIEAAVEGMVI
GCILPNGNYPGPKKEYKAVLPHRLTAYAAELLELDVPLVILAGDYNVMTLEDVYKP
ERNVNDALFRIEVDADVHRLLEOGWTDALRQLHGRVVTENDYFENAFARDAGLRID
HLLSPHVTIRLSAAGVDRHVRGWHTSDHAPWIELSDGPAEDQ"
5519. 5737
/gene="SMC00958"
CDS
5519. 5737
/gene="SMC00958"
/functions="miscellaneous; unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAC46229.1"
/db_xref="GI:15074584"
/db_xref="SPTREMBL:Q92PT0"
/translations="MTMNDWRAELRLLEKLVDAAVVAGARQDVAAIILDEIGQLRL
AVRDPDPADDISEQVIEEBSNNWPAAD"
5819. 6005
/notes="Sm-4 OR SMC04555
REPEAT SM-4
predicted by Homology"
/evidence=not_experimental
/complement(6137..9652)
/gene="mfd OR SMC04450"
/complement(6137..9652)
/gene="mfd OR SMC04450"
/functions="macromolecule metabolism; macromolecule
synthesis, modification; dna - replication, repair,
restr./modif"
/notes="Product confidence : probable
Gene name confidence : probable
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSCRIPTION-REPAIR COUPLING FACTOR
(TRCF) PROTEIN"
/protein_id="CAC46230.1"
/db_xref="GI:15074585"
/db_xref="SPTREMBL:Q92P89"
/translations="MTMIPGDERKILAAITREITIGVPTGAEALVLAEIARAGAPVA
YILSDGQKADLEQVGFAPDIPVLTLPWDCPLPYDRVSPSADTSARRLAALSALIA
HRRKPHAVLVTINAAQRISPDQVIESLAFTAPGNQIRMDDLAARLERNFERVP
TVRMGEFVARGGLDVPVPGSGEPLRLDFGDTLEAIRSFDPASORTTIGQVRSLDLN
PMSEVSLTPETIGHFRKOYLSLFGAATRDALYQAVSEGRRYAGMEHWLPLFYDRLET
VFDYLDGFRIYTDHIAPEAAERSKLVDYDARLASASPKSQVTCGTPYKPPEDM
LYLTAKGFGBALNDLNRLSPTEHEGEARQVNNIARGLRWAPAGAEUNDGRTI
NVFDQAVKHAERKARKAVIVSGWTEGSLDRLLQVLAHGLANIRPVKALSISLK
PGEAASVLEAGFTEGLDVIWIGEODILGDLRVRSRRKRKRGADFAEVTGLDEGSY
VWAHAGIGFVGLRTTEAGAPHDCLLELVADDAXLFLPVENIELLSRYGSEGTDAV
LDKLGVAQWAKARKKRLDLMAGGLIRIAAERHTHEAPVLAADQGVYDEPAARFPY
DEVEDQLNSIDAVRDLGRCPMDRLVCGVGFQKTEVALRAAFIAMNGVQVAVVP
TTLARQHTFFSDRFGLFIRIQOASRLVSGKDALTKKEVABGKTDIVVGHALLIG
SSIKFANGLLIIIDEEHFGVKHRELKELKLDVHVLTSATPIPTLQALTGVELRL
SLITTPVDMARVTFISPFDAVIRETLREHYRGQSFCVPCPRVSDLPETHDFLKS
DVPELKVAHQMPATELEDINNAFYEGRYDVLSTTIVESGLDVPNTANTLIVHRAAD
MFGLAQYLQLRGRVGRSKVAFALFTLPVNKLTGPAERRLKVLQSLDITLQAGFQLAS

```

```

HDLDIRGAGNLLGEEQSGHIKEVGFPELYQOMLEBAVELKGEERLHDTGWSPOISVGT
PVMIFEEYVPDLNLRGLYRRLGELTDLKEIDGFAELIDRFGLPTEVQHLKLIYV
Query Match 13.2%; Score 115; DB 1; Length 306250;
Best Local Similarity 53.6%; Pred. No. 1.8e-15;
Matches 309; Conservative 0; Mismatches 235; Indels 33; Gaps 2;
QY 286 AAACATGATGTTGATGGTTAGTTTTCGCGGGTTTGGAAATCGGTGCGCCCTTCCGCGCAAAATG 345
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297332 AAATCAGGCAAAATCAATGATTCAGAGCAATTCAGAGCTCGCGCGATTTGGCCCGACAGA 297391
QY 346 TTGATCGCGCGCTCGGCGGAATCTCATCGATCGCTCCAACTCGCGCTCAGAAACTCC 405
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297392 TTGACGTGCGCCCTCGCGGGCGTAAAGATCGATCTCGCTCAGTTCTCTCGGTGGAAGTG 297451
QY 406 AAGTTGTTGAGTGAATCAAGGGCTTTGTCCAGCTGTCTCAACTGACGAAGCAACCAATCAAT 465
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297452 TCGTTTTCAGCGCCTTGACGAGTCGACGATCTGTGCGGAACGGCTGGCGCGATCAGC 297511
QY 466 GCATCTGTCACGGTATTCGCGGCGGTACTCTCTTGTCTCGCGGACGACCCATGCAAGCGCC 525
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297512 GCCAGGTGATGCGCGCCG-----CGCGCAGCACCACGAGCGAGTGCC 297553
QY 526 ATCTGCGCAAGTCACTGCGCGGCTTCTCTGGCGATGTCATTGAGCTTTCGGGACCATATCA 585
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297554 AUCTGGCGAGCGTCTGGCGCGCGCTTTCGGCGATGGCGTTGAGTTCGCGCGCGGTGTCC 297613
QY 586 ATATTGTTCACTGTTCAACATGCGCTCAGACAGGAGCTTACCTCTGGCTGGCGCGGAAACCC 645
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297614 ATGATCCCCGCGACGGATGAAGTCTTCTTGGAGGAAGTGTCTGCGCGCGCGGTGTCC 297673
QY 646 TCTGGAATTCATTCGAGATATTGTCGCGAGCAGCGCCCTGCGCAAGTGGTGAGAAAGCA 705
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297674 TCGGGAATGCGGTTCAGGTACTTTCGTCGTCAGCATGCCCCCTGGCGCAGCGCGCGG 297733
QY 706 ATGACGCCAAGACCATTTGTCGCGAGCTGACTGCAACAAGTTCTCACCGTCATCGCCCGGT 765
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297734 ATGAGCCGACGCGCGAGTTCTCTCAGCGTGTGACGAGACCGTCA----- 297778
QY 766 TCCTCCACCCCAACGATTAATGATGGAATAGCTTGGCTGATGATCAGAAAGCGCGGAGCC 825
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297779 TCCTCGACCCCAACGTTGAGCATCGAATAGCTCGGCTGGTGGATGAGACAGGCGTGCC 297838
QY 826 TCCTCGCCCATGAACTCAGCGCGCTTCGGCTGTGAGCT 862
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297839 ATCTCTTCAGGATTACCGCGCGCTTCGCGGGTACGCT 297875
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: April 26, 2003, 23:44:41

Job time : 3110.77 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:02:03 ; Search time 1163.26 Seconds  
(without alignments)  
12154.292 Million cell updates/sec

Title: US-09-105-117K-1\_COPY\_1421\_2293

Perfect score: 873

Sequence: 1 gtcgataagcagcggttg.....ctgtgagctctgaccgtag 873

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	56.4	6.5	860	17 AF075981	AF075981 AF075981
C 2	48.8	5.6	1101	17 CNS017SY	AL108460 Drosophil
C 3	44.4	5.1	664	17 CNS03K4J	AL247708 Tetraodon
C 4	43.8	5.0	3237	17 BH770958	BH770958 L1Mtag68
C 5	41.6	4.8	558	14 BM712036	BM712036 UI-E-DWI-
C 6	41	4.7	617	14 BM729055	BM729055 UI-E-E01-

C 7	41	4.7	770	12 BE898495	BE898495 601681396
C 8	41	4.7	861	12 BE740880	BE740880 601593311
C 9	40	4.6	357	9 AI418751	AI418751 tg33a02.x
C 10	40	4.6	431	9 AA452858	AA452858 zx41C05.s
C 11	40	4.6	447	14 W91005	W91005 mfb3h05.r1
C 12	40	4.6	463	9 AA476500	AA476500 zx02h03.s
C 13	40	4.6	504	9 AI159901	AI159901 qb55h08.x
C 14	40	4.6	540	9 AI401605	AI401605 tf60b01.x
C 15	40	4.6	564	9 AI820014	AI820014 wj60a07.x
C 16	40	4.6	600	12 BG806166	BG806166 2021-07.M
C 17	39.8	4.6	1101	17 CNS00KK2	AL077673 Drosophil
C 18	39.6	4.5	925	17 CNS0091P	AL053013 Drosophil
C 19	39.2	4.5	846	17 CNS010RJ	AL093037 Drosophil
C 20	38.6	4.4	436	10 BE025218	BE025218 894008G12
C 21	38.6	4.4	513	14 BQ818660	BQ818660 1030072D0
C 22	38.6	4.4	828	17 CNS026XT	AL183962 Tetraodon
C 23	38.6	4.4	830	17 CNS01MQH	AL151258 Anopheles
C 24	38.6	4.4	1667	11 AY108197	AY108197 Zea mays
C 25	38.4	4.4	650	10 BB576857	BB576857 BB576857
C 26	38	4.4	212	9 AA150978	AA150978 z148c01.s
C 27	38	4.4	486	9 AA149286	AA149286 z125s05.s
C 28	37.8	4.3	804	17 CNS01HIJ	AL144476 Anopheles
C 29	37.4	4.3	707	10 AW190228	AW190228 x160e06.x
C 30	37.4	4.3	724	13 BI808515	BI808515 D002B11.O
C 31	37.4	4.3	731	14 BQ578715	BQ578715 WHE0308.H
C 32	37.2	4.3	448	17 BH614401	BH614401 1C22HF7.T
C 33	37.2	4.3	451	17 BH843100	BH843100 TC3-54P23
C 34	37.2	4.3	1623	17 BH614228	BH614228 1C22/2B8-
C 35	36.8	4.2	312	10 AW015929	AW015929 UI-H-B10P
C 36	36.8	4.2	482	12 BG048617	BG048617 OV1.14.H0
C 37	36.8	4.2	506	13 BI683498	BI683498 464507.MA
C 38	36.8	4.2	552	10 AW659102	AW659102 96030.WAR
C 39	36.8	4.2	692	17 CNS007WH	AL050323 Drosophil
C 40	36.6	4.2	636	13 BI724148	BI724148 1031069C0
C 41	36.6	4.2	819	12 BF064683	BF064683 HV.CEB001
C 42	36.6	4.2	837	13 BI868875	BI868875 603394947
C 43	36.4	4.2	660	13 BI328066	BI328066 602979206
C 44	36.4	4.2	910	17 CNS006ON	AL065629 Drosophil
C 45	36.4	4.2	934	12 BF311804	BF311804 601897316

#### ALIGNMENTS

RESULT 1  
LOCUS AF075981/c  
DEFINITION AF075981 Salmonella typhimurium LT2, Lambda DASH II Salmonella typhimurium genomic clone 390-T3, DNA sequence.  
ACCESSION AF075981  
VERSION AF075981.1 GI:3320851  
KEYWORDS GSS.  
SOURCE Salmonella typhimurium.  
ORGANISM Salmonella typhimurium.  
REFERENCE 1 (bases 1 to 860)  
AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.  
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library: comparison to the Escherichia coli K12 genome  
JOURNAL PEMS Microbiol. Lett. 173 (2), 411-423 (1999)  
MEDLINE 99243757  
COMMENT Contact: McClelland M  
Molecular Biology  
Sidney Kimmel Cancer Center  
3099 Science Park Road, San Diego, CA 92121, USA  
Email: mclelland@lifesci.sdsu.edu  
Class: shotgun.  
Location/Qualifiers  
1..860  
/organism="Salmonella typhimurium"  
/strain="LT2"  
/db\_xref="taxon:602"

```
/clone="390-T3"
/clone lib="Salmonella typhimurium LT2, Lambda DASH II"
/notes="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 212 a 199 c 242 g 207 t
ORIGIN
Query Match 6.5%; Score 56.4; DB 17; Length 860;
Best Local Similarity 53.5%; Pred. No. 0.00012;
Matches 199; Conservative 0; Mismatches 151; Indels 22; Gaps 3;
QY 505 CGCAGCACCATCAAGCCATCTCGCAAGTCACTGCGCGGTCTCTCGGGGATGTCA 564
DB 400 CGCAGCATCAGGCGCGCCATTTGATACATTTTGGACCCCGTCCGCGCAGCGATTCA 341
QY 565 TTGAGCTTGGCGACCATATCAATATTTGACGTTCAACATGCTCCAGCAGGAGCTTA 624
DB 340 TTCAACCGCGGAATTTTCCAGTTTTCGCGGGTATCTGTTCTGTTTAAGGAACGG 281
QY 625 CCCTGGCTG---GGCGGGAACCTCTGGAAATTCATCGAGATATTTGCTGGTGAAGG 681
DB 280 CTTCGCTCGCGCGCGGAATCTTCCGGAATACCAATTTCAATTTACGCTCAGTGAGC-TG 222
QY 682 CCCTGCCCAAGTGTGAGAAGCAATCAAGCAGCAAGACCATTTGTCGAGCTGACTGCAAC 741
DB 221 CCCGCCCCAGCGCGGAGAGCAATCTACCGACGCTTTTTCCTGCAACAGCGCCGAGC 162
QY 742 AAGTTCTCACCGTCATCGCCCGGTCTCTCCACCAACGATTAATGATGGAATAGCTTGGC 801
DB 161 AG-----CCCGTCTCTACCAACGCTCAAAAGCAATATTAGGC 120
QY 802 TGATGATCAGACGGGAGCGCTCTCTCGGCCATGAACCTCAGCGCTCGCTGTGAGC 861
DB 119 TGATGATCAGCAAGCGGTGCGGAGATCTCTCCAGATATATCGATAGCTTCTGTGGCAGA 60
QY 862 TCTGACCGTAG 873
DB 59 TCGCAGGATAG 48
RESULT 2
CNS017SV/c 1101 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BAC37L08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108460
VERSION AL108460.1 GI:5628764
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
TITLE 1 (bases 1 to 1101)
JOURNAL Genoscope.
COMMENT Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaut at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelOBAC11.
FEATURES
Location/Qualifiers
1. .1101
/organisms="Drosophila melanogaster"
/db xref="taxon:7227"
/clone="BACN37L08"
```

```
/clone lib="DrosBAC"
/plasmid="pBelOBAC11"
/notes="end : SP6"
BASE COUNT 254 a 176 c 160 g 152 t 359 others
ORIGIN
Query Match 5.6%; Score 48.8; DB 17; Length 1101;
Best Local Similarity 14.5%; Pred. No. 0.018;
Matches 47; Conservative 156; Mismatches 121; Indels 0; Gaps 0;
QY 72 GTATTGGACGCGTGTGCTTTATCGCGCGTCCGCGCAATACGCGACACCGGACG 131
DB 1079 KTTTBTBTSSSTSTSTBTBSSSSBBSSSSSSBSSSTSSSSSTSSSS 1020
QY 132 GTGAGTTTTCGCGTGGCGCGTTCGCGCAAGCTGATCTGTTCCCTCGCTGGGTTT 191
DB 1019 SSTSBTBTSSSTSTSTTTTBTBSSBSSSSSTSSSBTSBTSBSS 960
QY 192 CGCGCAGCAGCATTTCTACGCGCGCTGCCAGCCCAAGTGTGGCTGATCAACGT 251
DB 959 BSTSSASBSSSSSSBSSSTTTBTSBSSSTSSSSSSSSBBSBTSBTSBTTT 900
QY 252 CGTCTGGCAGTTGTGATGACGCGATTGCGCATCAACTGATGTTGATGGTTAGTTTC 311
DB 899 BTXSTSTSSSTSBTTTSSSSSSSSSSSSBSSSTTSBSSBSSBTSBTSBTSB 840
QY 312 GCGGGTTTGGATCGGTGCGCTTCGCCCAATTTGATGCGCGCGCTCGTGGAAATCTC 371
DB 839 TSGSTBTSTBMSKBSSTSTSTBTSSGCGGCGSTSGSCSBCGKSTSSGTSSTGTT 780
QY 372 ATCGATCGCTCAACTCGCGTC 395
DB 779 GBCSTGGCSCCCYCCCTCSTC 756
RESULT 3
CNS03K4J 664 bp DNA linear GSS 17-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 032B06 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL247708
VERSION AL247708.1 GI:7968720
KEYWORDS GSS; Genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
REFERENCE 1 (bases 1 to 664)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
TITLE 1 (bases 1 to 664)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
JOURNAL Human gene number estimate provided by genome wide analysis using
REFERENCE Tetraodon nigroviridis DNA sequence
AUTHORS Unpublished
TITLE 2 (bases 1 to 664)
AUTHORS Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
JOURNAL Characterization and repeat analysis of the compact genome of the
REFERENCE freshwater pufferfish Tetraodon nigroviridis
AUTHORS 3 (bases 1 to 664)
TITLE Genoscope.
JOURNAL Direct Submission
COMMENT Submitted (12-APR-2000)
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES Location/Qualifiers
```

```

source
1..664
/organism="Tetradodon nigroviridis"
/db_xref="taxon:99883"
/clone_lib="G"
/clone_lib="G"
/note="Genoscope sequence ID : COBG032DA03LP1-end : T7"

BASE COUNT      118 a   200 c   222 g   116 t      8 others
ORIGIN

Query Match      5.1%; Score 44.4; DB 17; Length 664;
Best Local Similarity 48.2%; Pred. No. 0.23;
Matches 120; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 181 CTGCTGGGTTTCGGCGCAGCAGCATTTGTCACGCGCTGTTCAGGCCCCCAAGTGTGGCGC 240
Db 278 CAGGTCCGCTCTGTCGCCGGAATATGCGGGGCGCCCGCGGACGAGCCGCGCGANGCAC 337
Qy 241 TGGATCAACGTCGTCTGGCAGTTCTGTGATGACCGCATTTGGCCCATCAAACTCATGTTGATG 300
Db 338 CGCGGNTCTTTTGTGTCGGCGCGTGCACGGGACAGGCCCGGCGACCGGCGCTGNTGTGTGTG 397
Qy 301 GGTATGTTTTCCGGGTTTTGGAATCGGTGGGCTTTGCGCCAAATGTTGATGCCGCGCGTCG 360
Db 398 GCGAACGATCCGTTGGGTGCGGASCCGCTGCTCTCGCACGATCGAGAGGCCCTGGGGGCG 457
Qy 361 TGGGAAATCTCATCGATCGCTCCAACTCGGGGTGAGAAATCTCAAGTGTGTTAGTGAA 420
Db 458 GAGGCTCCCCCACCGTGGARCTCCCCGCGCGTGGAGGCTCCCCCGTGGTGGGGTCTT 517
Qy 421 TCAAGGCTG 429
Db 518 CCGCGCGC 526

RESULT 4
BH770958
LOCUS      BH770958          3237 bp      DNA      linear      GSS 01-MAY-2002
DEFINITION LMGtag686 MG1363 Random Sequence Tag Library Lactococcus lactis
            subsp. cremoris genomic, DNA sequence.
VERSION     BH770958
KEYWORDS    BH770958.1 GI:20373915
SOURCE      GSS.
ORGANISM    Lactococcus lactis subsp. cremoris.
            Lactococcus lactis subsp. cremoris
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Lactococcus.
REFERENCE   1 (bases 1 to 3237)
AUTHORS    Bolotin,A., Ehrlich,S.D. and Sorokin,A.
TITLE      Studies of genomes of dairy bacteria Lactococcus lactis
JOURNAL    Sci. Alimente, (2002) In press
COMMENT    Contact: Sorokin A
            Genetique Microbienne
            INRA
CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
Tel: 33 1 34 65 25 16
Fax: 33 1 34 65 25 21
Email: sorokine@jouy.inra.fr
Class: shotgun
High quality sequence start: 30
High quality sequence stop: 3207.
Location/Qualifiers
1..3237
/organism="Lactococcus lactis subsp. cremoris"
/strain="MG1363"
/db_xref="taxon:1359"
/clone_lib="MG1363 Random Sequence Tag Library"
/note="Vector: PSGM2; Site1: SmaI; Library of
chromosomal fragments of L.lactis strain MG1363 was
prepared by partial AluI digestion or by sonication."

BASE COUNT      1058 a   577 c   613 g   980 t      9 others
ORIGIN

```



Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLC819 row: p column: 23  
 High quality sequence stop: 765.

# FEATURES

Source  
 1. 770  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3951622"  
 /clone\_lib="NIH MGC 9"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 147 a 245 c 277 g 101 t

Query Match 4.7%; Score 41; DB 12; Length 770;  
 Best Local Similarity 46.9%; Pred. No. 2.2;  
 Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
 QY 98 GCGCGTGGCGCGCAATACGCGACACCGGAGTTCGCGCGTGGCGCGTTCG 157  
 |||||  
 Db 385 GCAGCGCAGCGCGCGAATGATCTCCAGTCGCGGAACTCGCGTTCGCGTCA 326  
 QY 158 CGGAAGCTGATGTTCCCGTGGTGGTTCGCGCAGCAGCATTTGTCAGCCCGC 217  
 |||||  
 Db 325 CCACCGTGGTGGATGCGACACCGCTGATGTCGCGTTCGCGCGCTTGGCGCGC 266  
 QY 218 TGTCCAGCCCCAGGTGGCGTGGATCAACGTCGTCGCGGATGTCGATGACCCAT 277  
 |||||  
 Db 265 TGTCAACCGCAGCGTCCCGAGTGGTCTCGGCGGTGGTCTTGAAGTAGTGGC 206  
 QY 278 TGGCCATCAAACTGATGTTGATGGTGTAGTTTTCGCGGGTTTGGATCGTGGCTTCG 337  
 |||||  
 Db 205 TGTCTTGGCTCGATGTTGAGTGCAGTCTCCAGTAGAAGCGTGTGTAGCAGG 146  
 QY 338 CCCAAATGTTGATCGCGCGTGGTGGAAATCT 370  
 |||||  
 Db 145 CCGCCACTTGTGATGATGCTCGTGGTGGATGT 113

RESULT 8  
 BE740880/c  
 LOCUS  
 DEFINITION 601593311F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3947138 5', mRNA linear EST 15-SBP-2000  
 mRNA sequence.  
 ACCESSION BE740880  
 VERSION BE740880  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 861)  
 AUTHORS NIH-MGC http://imgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov  
 Plate: LLC808 row: f column: 03  
 High quality sequence stop: 754.

# FEATURES

Source  
 1. 861  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3947138"  
 /clone\_lib="NIH MGC 9"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: ovary; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 187 a 245 c 285 g 144 t

Query Match 4.7%; Score 41; DB 12; Length 861;  
 Best Local Similarity 46.9%; Pred. No. 2.3;  
 Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;  
 QY 98 GCGCGTGGCGCGCAATACGCGACACCGGAGTTCGCGCGTGGCGCGTTCG 157  
 |||||  
 Db 675 GCAGCGCAGCGCGCGAATGATCTCCAGTCGCGGAACTCGCGTTCGCGTCA 616  
 QY 158 CGGAAGCTGATGTTCCCGTGGTGGTTCGCGCAGCAGCATTTGTCAGCCCGC 217  
 |||||  
 Db 615 CCACCGTGGTGGATGCGACACCGCTGATGTCGCGTTCGCGCGCTTGGCGCGC 556  
 QY 218 TGTCCAGCCCCAGGTGGCGTGGATCAACGTCGTCGCGGATGTCGATGACCCAT 277  
 |||||  
 Db 555 TGTCAACCGCAGCGTCCCGAGTGGTCTCGGCGGTGGTCTTGAAGTAGTGGC 496  
 QY 278 TGGCCATCAAACTGATGTTGATGGTGTAGTTTTCGCGGGTTTGGAACTCGTGGCTTCG 337  
 |||||  
 Db 495 TGTCTTGGCTCGATGTTGAGTGCAGTCTCCAGTAGAAGCGTGTGTAGCAGC 436  
 QY 338 CCCAAATGTTGATCGCGCGTGGTGGAAATCT 370  
 |||||  
 Db 435 CCGCCACTTGTGATGATGCTCGTGGTGGATGT 403

RESULT 9  
 AI418751  
 LOCUS  
 DEFINITION tg33a02.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2110538 3', mRNA linear EST 30-MAR-1999  
 mRNA sequence.  
 ACCESSION AI418751  
 VERSION AI418751.1 GI:4264682  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 357)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.





cDNA Library Arrayed by: Greg Lennon, Ph. D.  
 DNA Sequencing by: Washington University Genome  
 Clone distribution: NCI-CGAP clone distribution  
 found through the I. M. A. E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Insert Length: 671  
 Seq primer: -40ml3 fwd, ET from Amersham  
 High quality sequence stop: 482.









## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	873	100.0	2374	18 AAT56816	DNA encoding LysG,
C 2	873	100.0	34980	22 AAH68528	C glutamicum codin
C 3	568	65.1	627	22 AAH68419	C glutamicum codin
4	404.2	46.3	993	22 AAF68077	Corynebacterium gl
5	326	37.3	822	22 AA71779	Corynebacterium gl
6	326	37.3	822	23 AAS96098	C. glutamicum gene
7	303	34.7	708	22 AAH68420	C. glutamicum codin
8	193.6	22.2	1568	22 AAH45375	C. thermoaminogene
9	71.6	8.2	5541	15 AAQ55755	Escherichia coli g

## ALIGNMENTS

PF 22-DEC-1995; 95DE-1048222

PF 22-DEC-1995; 95DE-1048222

PR 22-DEC-1995; 95DE-1048222.  
 XX (KERJ ) FORSCHUNGSZENTRUM JUEBLICH GMBH.  
 XX Egging L, Sahm H, Vrljic M;  
 XX WPI; 1997-333867/31.  
 DR P-PSDB; AAW37714-16.  
 XX  
 XX Increasing microbial production of amino acids, especially lysine -  
 PT by improving export carrier activity or corresponding gene  
 PT expression, also new export and regulatory genes from  
 PT Corynebacterium  
 XX  
 PS Claim 23 and 26; Page -; 16pp; German.  
 XX  
 CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,  
 CC LysE and ORE3 genes. LysG and LysE encode a lysine transport regulatory  
 CC protein and an export protein, respectively. Microbial production of  
 CC amino acids (A) is improved by increasing the export-carrier activity  
 CC and/or the export gene expression in a microorganism that produces (A).  
 CC The method is specifically used to increase production of lysine, as  
 CC used as an animal feed additive. Other (A) are variously useful as  
 CC pharmaceuticals, condiments and intermediates for fine chemicals.  
 CC This method increases the amount of (A) secreted into the culture medium.  
 CC Export of (A) has been found to depend on a single gene.  
 CC NB: This sequence has been created from the information given in table 2  
 CC of the specification.  
 XX  
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

Query Match 100.0%; Score 873; DB 18; Length 2374;  
 Best Local Similarity 100.0%; Pred. No. 7.3e-247;  
 Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTGCACCTGGTGG 60  
 DB 1421 GTCCATAAGCAGCGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTGCACCTGGTGG 1480  
 QY 61 AACCCGAATGCGTATTTGGAGCGGTTTGTGTTATCGCGCGCGTCCGGCGCAATACGGC 120  
 DB 1481 AACCCGAATGCGTATTTGGAGCGGTTTGTGTTATCGCGCGCGTCCGGCGCAATACGGC 1540  
 QY 121 GACACCGAGCGGTGGATTTTCGCGCTGCGCGCTTTCGCGGCAAGCCTGATCTGTTCCCG 180  
 DB 1541 GACACCGAGCGGTGGATTTTCGCGCTGCGCGCTTTCGCGGCAAGCCTGATCTGTTCCCG 1600  
 QY 181 CTGGTGGGTTTTCGCGCGAGCAGCATTTGTCAGCGCGCTGTCAGCGCCCAAGGTGTGGCG 240  
 DB 1601 CTGGTGGGTTTTCGCGCGAGCAGCATTTGTCAGCGCGCTGTCAGCGCCCAAGGTGTGGCG 1660  
 QY 241 TGGATCAACGTCGTCGTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG 300  
 DB 1661 TGGATCAACGTCGTCGTGGCAGTTGTGATGACCGCATTTGGCCATCAAACTGATGTTGATG 1720  
 QY 301 GGTAGTTTTCGCGGTTTGGATCGTGGCTTTCGCGCCAAATGTTGATCGCGCGTGG 360  
 DB 1721 GGTAGTTTTCGCGGTTTGGATCGTGGCTTTCGCGCCAAATGTTGATCGCGCGTGG 1780  
 QY 361 TGGGAATCTCATGTCGCTCCAACTCGCGCTCAGAAACTCCAACTGTTGTTGAGTGAA 420  
 DB 1781 TGGGAATCTCATGTCGCTCCAACTCGCGCTCAGAAACTCCAACTGTTGTTGAGTGAA 1840  
 QY 421 TCAAGGCTGTTTCCAGCTGCTCAACTGACGAAGCACCACCAATGCACTGGTCAACGGTA 480  
 DB 1841 TCAAGGCTGTTTCCAGCTGCTCAACTGACGAAGCACCACCAATGCACTGGTCAACGGTA 1900  
 QY 481 TCCGCGCGTACTCTCTTCTGCGGAGCACCACCAATGCAAGCGGCATCTGCGCAAGTGAC 540  
 DB 1901 TCCGCGCGTACTCTCTTCTGCGGAGCACCACCAATGCAAGCGGCATCTGCGCAAGTGAC 1960  
 QY 541 TCCGCGCGTCTCTCTGCGGAGTCTATTGAGCTTTCGCGACCATATCAATATTCTTCACGTT 600

DB 1961 TGCCCGGTTCTGGCGGATGTCAATTGAGCTTGGCGACCATATCAATATGTTTACGTTTC 2020  
 QY 601 AACATGCCCTCAGACAGGAGACTTACCTTGGCTGGCGGGGAAACCTCTCTGGAATTCATCG 660  
 DB 2021 AACATGCCCTCAGACAGGAGACTTACCTTGGCTGGCGGGGAAACCTCTCTGGAATTCATCG 2080  
 QY 661 AGATATTTGTTCGTGAGCAGGCGCTTCGCGCAAGTGTGTGAGAAAGCAATGACGCCAAGACCA 720  
 DB 2081 AGATATTTGTTCGTGAGCAGGCGCTTCGCGCAAGTGTGTGAGAAAGCAATGACGCCAAGACCA 2140  
 QY 721 TTGTTGCGAGCTGACTGCAACAAAGTTTCTCACGTCATCGCCCGGTTCTCCACCCAAACGA 780  
 DB 2141 TTGTTGCGAGCTGACTGCAACAAAGTTTCTCACGTCATCGCCCGGTTCTCCACCCAAACGA 2200  
 QY 781 TTAATGATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCCGCAATGAAC 840  
 DB 2201 TTAATGATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCTCCGCAATGAAC 2260  
 QY 841 TCAGCCGCTCTCGCTGTGAGCTCTGACCCGTAG 873  
 DB 2261 TCAGCCGCTCTCGCTGTGAGCTCTGACCCGTAG 2293

## RESULT 2

AAH68528/c  
 ID AAH68528 standard; DNA; 349980 BP.

XX AC AAH68528;

XX DT 26-SEP-2001 (first entry)

XX DE C glutamicum coding sequence fragment SEQ ID NO: 7063.

XX KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX OS organic acid synthesis; ds.

XX OS Corynebacterium glutamicum.

XX PN BP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX DR WPI; 2001-376931/40.

XX PT Novel polynucleotides derived from Coryneform bacteria, for identifying mutation point of a gene, measuring expression of a gene, analysing expression profile or pattern of a gene and identifying homologous gene

XX PS Disclosure; SEQ ID NO: 7063; 246pp + Sequence Listing; English.

XX CC The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

XX CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX  
SQ Sequence 349980 BP; 80727 A; 91049 C; 97618 G; 80586 T; 0 other;  
Query Match 100.0%; Score 873; DB 22; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 6.1e-246;  
Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG 60  
DB 128548 GTCGATAGCAGCGGTTTGGGTAAGCCCATGTTGATGGCAATCGTCTGACCTGGTTG 128489  
QY 61 AACCCGAATGCGTATTGGAACGCGTTGTTATCGCGCGCGTGGCGCAATACGGC 120  
DB 128488 AACCCGAATGCGTATTGGAACGCGTTGTTATCGCGCGCGTGGCGCAATACGGC 128429  
QY 121 GACACCGGAGCGTGGATTTCGCGCGTGGCGGTTTCGCGGCAAGCTGATCTGGTCCG 180  
DB 128428 GACACCGGAGCGTGGATTTCGCGCGTGGCGGTTTCGCGGCAAGCTGATCTGGTCCG 128369  
QY 181 CTGTGGGTTTCGCGCGAGCAGCATTTGTACGCGCGCTGTCCAGCCCAAGGTGTGCGC 240  
DB 128368 CTGTGGGTTTCGCGCGAGCAGCATTTGTACGCGCGCTGTCCAGCCCAAGGTGTGCGC 128309  
QY 241 TGGATCAACGTCGTGCGGAGTTGTGATGACCGCATTTGGCCATCAAACTGATTTGATG 300  
DB 128308 TGGATCAACGTCGTGCGGAGTTGTGATGACCGCATTTGGCCATCAAACTGATTTGATG 128249  
QY 301 GGTAGTTTTCGCGGTTTGGATCGGTGGCTTCGCCCAATGTTGATGCGCGGCTCG 360  
DB 128248 GGTAGTTTTCGCGGTTTGGATCGGTGGCTTCGCCCAATGTTGATGCGCGGCTCG 128189  
QY 361 TGGGAATCTCATCGATCGCTCCAACTCGCGGTCAGAACTCCAAAGTTGTCAGTGAA 420  
DB 128188 TGGGAATCTCATCGATCGCTCCAACTCGCGGTCAGAACTCCAAAGTTGTCAGTGAA 128129  
QY 421 TCAAGGCTGTGTCCAGTGTCTCAACTGAGCAGCAACCAATCAATGCACTGGTCACGGTA 480  
DB 128128 TCAAGGCTGTGTCCAGTGTCTCAACTGAGCAGCAACCAATCAATGCACTGGTCACGGTA 128069  
QY 481 TCGCGCGGCTACTCTCTCTGTCGCGGAGCAGCACCATGAGCGGCTCTGCGCAAGTGAC 540  
DB 128068 TCGCGCGGCTACTCTCTCTGTCGCGGAGCAGCACCATGAGCGGCTCTGCGCAAGTGAC 128009  
QY 541 TGCCCGGTTCTCGGCGATGTCAATGAGCTTGGGACCATATCAATATTTTCACTTC 600  
DB 128008 TGCCCGGTTCTCGGCGATGTCAATGAGCTTGGGACCATATCAATATTTTCACTTC 127949  
QY 601 AACATGCCCTCAGACAGGACTTACCTCGTGGCGGGGAAACCTCTGGAATTTCCATCG 660  
DB 127948 AACATGCCCTCAGACAGGACTTACCTCGTGGCGGGGAAACCTCTGGAATTTCCATCG 127889  
QY 661 AGATATTTGTCGTGACAGCGCTTCGCGAAGTGGTGAGAAAGCAATGACCAAGACCA 720  
DB 127888 AGATATTTGTCGTGACAGCGCTTCGCGAAGTGGTGAGAAAGCAATGACCAAGACCA 127829  
QY 721 TTGTTGACGCTGACTCCAAAGTTCTCACCGTCATCGCGCGGTTCTCCACCAACGA 780  
DB 127828 TTGTTGACGCTGACTCCAAAGTTCTCACCGTCATCGCGCGGTTCTCCACCAACGA 127769  
QY 781 TTAATGATGGAATAGCTTGGCTGATGAATCAGAAAGCGGCGAGCCCTCTCCGCGCATGAAC 840  
DB 127768 TTAATGATGGAATAGCTTGGCTGATGAATCAGAAAGCGGCGAGCCCTCTCCGCGCATGAAC 127709  
QY 841 TCAGCGCCCTCCGCTGAGCTCTGGACCGTAG 873  
DB 127708 TCAGCGCCCTCCGCTGAGCTCTGGACCGTAG 127676

RESULT 3  
AAH68419/c  
ID AAH68419 standard; DNA; 627 BP.

XX  
AC

AAH68419;

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 3454.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis; ds.

OS Corynebacterium glutamicum.

XX BP1108790-A2.

PN 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

DR WPI: 2001-376931/40.

DR P-PSDB; AAG93200.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 1; SEQ ID NO: 3454; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from corynebacterium, and identifying a homologue of a gene derived  
CC from corynebacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX SQ Sequence 627 BP; 139 A; 159 C; 185 G; 144 T; 0 other;

Query Match 65.1%; Score 568; DB 22; Length 627;  
Best Local Similarity 100.0%; Pred. No. 3.4e-157;  
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 GTTTTCGCGGGTTTGGAAATCGGTGGCTTCGCCCAATGTTGATGCGCGGCTCGTGGGA 365

DB 627 GTTTTCGCGGGTTTGGAAATCGGTGGCTTCGCCCAATGTTGATGCGCGGCTCGTGGGA 568

QY 366 AATCTCATGATCGGCTCCACTCGGCGTCAGAAATCCCAAGTTGTTGAGTGAATCAAG 425

DB 567 AATCTCATGATCGGCTCCACTCGGCGTCAGAAATCCCAAGTTGTTGAGTGAATCAAG 508

QY 426 GCTGTTGTCAGCTGCTCAACTGACGAAGACCAATCAATGCACTGGTCACGGTATCGC 485

DB 507 GCTGTTGTCAGCTGCTCAACTGACGAAGACCAATCAATGCACTGGTCACGGTATCGC 448

QY 486 GCGGTACTCTCTGCTCGCGCAGCACCCATGCAAGCGCCATCTGCGCAAGTGAATCGCC 545

DB 447 GCGGTACTCTCTGCTCGCGCAGCACCCATGCAAGCGCCATCTGCGCAAGTGAATCGCC 388

QY 546 GCGTTCCTGCGGATGTCATTGAGCTTGGCGGACCATATCAATATTTGTTACGTTCAACAT 605  
DB 387 GCGTTCCTGCGGATGTCATTGAGCTTGGCGGACCATATCAATATTTGTTACGTTCAACAT 328  
QY 606 GCCCTCAGACGAGGACTTACCTGGCTGGCGGCGGAAACCTCTCTGGAATTCATCGAGATA 665  
DB 327 GCCCTCAGACGAGGACTTACCTGGCTGGCGGCGGAAACCTCTCTGGAATTCATCGAGATA 268  
QY 666 TTTGTTCGTGAGCAGGCGCTCGCGAAGTGTGAGAAAGCAATGACGCGCAAGACATTTT 725  
DB 267 TTTGTTCGTGAGCAGGCGCTCGCGAAGTGTGAGAAAGCAATGACGCGCAAGACATTTT 208  
QY 726 GGCAGCTGACTGCAACAAGTTCTCACGTCATCGCCGCTTCTCCACCCCAAGATTAAT 785  
DB 207 GGCAGCTGACTGCAACAAGTTCTCACGTCATCGCCGCTTCTCCACCCCAAGATTAAT 148  
QY 786 GATGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCCAATGAACCTCAGC 845  
DB 147 GATGAATAGCTTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCCAATGAACCTCAGC 88  
QY 846 GGCCTCCGCTGAGCTCTGAGACCGTAG 873  
DB 87 GGCCTCCGCTGAGCTCTGAGACCGTAG 60

RESULT 4  
AAF68077/c  
ID AAF68077 standard; DNA; 993 BP.  
AC AAF68077;  
XX  
DT 11-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:669.  
XX  
KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
KW genome mapping; genetic engineering; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN W0200100805-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000MO-IB00926.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031454.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031563.  
PR 09-JUL-1999; 99DE-1032122.  
PR 09-JUL-1999; 99DE-1032124.  
PR 09-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032128.  
PR 09-JUL-1999; 99DE-1032180.  
PR 09-JUL-1999; 99DE-1032182.  
PR 09-JUL-1999; 99DE-1032190.  
PR 09-JUL-1999; 99DE-1032191.  
PR 09-JUL-1999; 99DE-1032209.  
PR 09-JUL-1999; 99DE-1032212.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032927.  
PR 14-JUL-1999; 99DE-1033005.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040767.

PR 27-AUG-1999; 99DE-1040830.  
PR 27-AUG-1999; 99DE-1040831.  
PR 27-AUG-1999; 99DE-1040832.  
PR 27-AUG-1999; 99DE-1040833.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041395.  
PR 31-AUG-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042078.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042088.  
XX (BADI ) BASF AG.  
XX  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
PI P-PSDB; AAB76844.  
XX  
DR WPI; 2001-071486/08.  
DR  
XX P-PSDB; AAB76844.  
XX  
PT Corynebacterium glutamicum nucleic acids encoding membrane construction  
PT and membrane transport proteins or their portions, useful for typing or  
PT identifying C. glutamicum or related bacteria, and as markers for  
PT transformation .  
XX  
PS Claim 3; Page 1107-1108; 1119pp; English.  
XX  
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
CC construction and membrane transport (MCT) proteins given in AAB76510 to  
CC AAB76847. The MCT nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
CC identification of C. glutamicum or related bacteria, as reference points  
CC for mapping C. glutamicum genome, and as markers for transformation.  
CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
CC example from the present invention.  
XX  
SQ Sequence 993 BP; 222 A; 247 C; 277 G; 247 T; 0 other;  
XX  
Query Match 46.3%; Score 404.2; DB 22; Length 993;  
Best Local Similarity 98.8%; Pred. No. 7.9e-109;  
Matches 418; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 453 AGCACCAATCAATGACTGTGTCACGGTA--TCGGCGCGTACTCTCTCTGCTCGCGCAGC 510  
DB 993 AGCACCAATCAATGACTGTGTCACGGTAATCCGCGCGCGTACTCTCTCTGCTCGCGCAGC 934  
QY 511 ACCATGCAAGCGCCATCTGCGCAAGTACTGCCCGGTTCTTGGCGGATGTCATTGAGC 570  
DB 933 ACCATGCAAGCGCCATCTGCGCAAGTACTGCCCGGTTCTTGGCGGATGTCATTGAGC 874  
QY 571 TTGCGGACCATATCAATATTTGTTCAAGTTCAACATGCCCTCAGACAGGACTTACCCCTGG 630  
DB 873 TTGCGGACCATATCAATATTTGTTCAAGTTCAACATGCCCTCAGACAGGACTTACCCCTGG 814  
QY 631 CTGGCGCGGGAACCTCTCTGGAATTCATCGAGATTTTCTCGTGAGCAGGCGCTCGCGCA 690  
DB 813 CTGGCGCGGGAACCTCTCTGGAATTCATCGAGATTTTCTCGTGAGCAGGCGCTCGCGCA 754  
QY 691 AGTGGTGAGAAAGCAATGACGCGCAAGACCATTTGTTGGCAGCTGACTGCAACAAAGTTCTCA 750  
DB 753 AGTGGTGAGAAAGCAATGACGCGCAAGACCATTTGTTGGCAGCTGACTGCAACAAAGTTCTCA 694  
QY 751 CGGTCAATCGCCGGTTCTCTCCACCCCAAGATTAATGATGAATAGCTTGGCTGATGAATC 810  
DB 693 CGGTCAATCGCCGGTTCTCTCCACCCCAAGATTAATGATGAATAGCTTGGCTGATGAATC 634  
QY 811 AGAAGCGGCGAGCCCTCTCTCGGCATGAATCAGCGCGCTCGCTGTGAGCTCTGAGCCG 870  
DB 633 AGAAGCGGCGAGCCCTCTCTCGGCATGAATCAGCGCGCTCGCTGTGAGCTCTGAGCCG 574  
QY 871 TAG 873  
|||

Db 573 TAG 571

RESULT 5  
AAAF71779  
ID AA71779 standard; DNA; 822 BP.  
XX  
AC AA71779;  
XX  
DT 30-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:53.  
XX  
KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
KW fine chemical production; microorganism; organic acid; nucleoside;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
FN WO200100843-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-IB00923.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 01-JUL-1999; 99DE-1030476.  
PR 02-JUL-1999; 99US-0142101.  
PR 08-JUL-1999; 99DE-1031415.  
PR 08-JUL-1999; 99DE-1031418.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031435.  
PR 08-JUL-1999; 99DE-1031443.  
PR 08-JUL-1999; 99DE-1031453.  
PR 08-JUL-1999; 99DE-1031457.  
PR 08-JUL-1999; 99DE-1031465.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031541.  
PR 08-JUL-1999; 99DE-1031573.  
PR 08-JUL-1999; 99DE-1031592.  
PR 08-JUL-1999; 99DE-1031632.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1031636.  
PR 08-JUL-1999; 99DE-1032125.  
PR 09-JUL-1999; 99DE-1032126.  
PR 09-JUL-1999; 99DE-1032130.  
PR 09-JUL-1999; 99DE-1032186.  
PR 09-JUL-1999; 99DE-1032206.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032228.  
PR 09-JUL-1999; 99DE-1032229.  
PR 09-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032922.  
PR 14-JUL-1999; 99DE-1032926.  
PR 14-JUL-1999; 99DE-1032928.  
PR 14-JUL-1999; 99DE-1033004.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 12-AUG-1999; 99US-0148613.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.  
PR 27-AUG-1999; 99DE-1040832.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041380.

PR 31-AUG-1999; 99DE-1041394.  
PR 31-AUG-1999; 99DE-1041396.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042124.  
PR 03-SEP-1999; 99DE-1042129.  
PR 09-MAR-2000; 2000US-0187970.  
XX  
XX (BADI ) BASF AG.  
PA  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
PI WPI; 2001-137957/14.  
XX  
DR P-PSDB; AAB79660.  
XX  
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic  
PT pathway proteins, useful for producing fine chemicals in  
PT microorganisms, including organic acids, nonproteinogenic amino acids,  
PT and purine and pyrimidine bases -  
XX  
XX Claim 3; Page 233-234; 1737pp; English.  
PS  
XX AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
CC MP nucleic acids are useful for the production of fine chemicals  
CC in microorganisms, including organic acids, nonproteinogenic amino  
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
CC compounds, vitamins, cofactors, polyketides and enzymes.  
XX  
SQ Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;

Query Match 37.3%; Score 326; DB 22; Length 822;  
Best Local Similarity 100.0%; Pred. No. 8.2e-86;  
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGGTTG 60  
Db 497 GTCGATAAGCAGCGGGTTTGGGTAAAGCCCATGTTGATGGCAATCGTGTGACCTGGTTG 556  
QY 61 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGGTGCGCGCAATACGCG 120  
Db 557 AACCCGAATGCGTATTGGACGCGTTTGTGTTATCGCGCGGTGCGCGCAATACGCG 616  
QY 121 GACACCGGACGCGTGGATTTTCGCGCGGTGCGCGCAAGCTGATCTGTTCCCG 180  
Db 617 GACACCGGACGCGTGGATTTTCGCGCGGTGCGCGCAAGCTGATCTGTTCCCG 676  
QY 181 CTGGTGGGTTTCGCGCGCAGCAGCATTTGTCACGCCCGCTGCCAGCCCAAGGTGGCGC 240  
Db 677 CTGGTGGGTTTCGCGCGCAGCAGCATTTGTCACGCCCGCTGCCAGCCCAAGGTGGCGC 736  
QY 241 TGGATCAACGTCGTGTGCGCAGTTGATGACCGCATGGGCCATCAACTGATGTTGATG 300  
Db 737 TGGATCAACGTCGTGTGCGCAGTTGATGACCGCATGGGCCATCAACTGATGTTGATG 796  
QY 301 GGTTAGTTTTCGCGCGGTTTGGAAATC 326  
Db 797 GGTTAGTTTTCGCGCGGTTTGGAAATC 822

RESULT 6  
AAS96098  
ID AAS96098 standard; DNA; 822 BP.  
XX  
AC AAS96098;  
XX  
DT 26-FEB-2002 (first entry)





QY 181 CTGTTGGGTTTTCGGCGCAGCAGCATTTGTCACGCCCGCTGTCCAGCCCAAGGTGTGGCGC 240  
DB 586 CTGTTGGGTTTTCGGCGCAGCAGCATTTGTCACGCCCGCTGTCCAGCCCAAGGTGTGGCGC 645  
QY 241 TGGATCAACGTCGTCGGCGAGTTGTGATGACCGCATTTGGCCATCAAACTGATTTGATG 300  
DB 646 TGGATCAACGTCGTCGGCGAGTTGTGATGACCGCATTTGGCCATCAAACTGATTTGATG 705  
QY 301 GGT 303  
DB 706 GGT 708  
RESULT 8  
AAH45375  
ID AAH45375 standard; DNA; 1568 BP.  
XX AC AAH45375;  
XX DT 11-SEP-2001 (first entry)  
XX C. thermoaminogenes lysin biosynthetic enzyme lysE DNA.  
XX DE Heat-resistant; lysin biosynthesis; enzyme; coryneform;  
XX KW aspartate-semialdehyde dehydrogenase; lysE; ds.  
XX OS Corynebacterium thermoaminogenes.  
XX PN JP2001120270-A.  
XX PD 08-MAY-2001.  
XX PF 01-NOV-1999; 99JP-03111148.  
XX PR 01-NOV-1999; 99JP-03111148.  
XX PS (AJIN ) AJINOMOTO KK.  
XX PT WPI; 2001-364760/38.  
XX DR P-PSDB; AAG64047.  
XX A heat-resistant lysin biosynthetic system enzyme gene of a high  
PT temperature-resistant coryneform microbe -  
XX Example 5; Page 22-24; 27pp; Japanese.  
XX The invention relates to a gene from a high temperature-resistant  
CC coryneform microbe that encodes a heat-resistant lysin biosynthetic  
CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity  
CC and can be used for growing amino acid-producing microbes. The  
CC present sequence encodes an enzyme of the invention.  
XX SQ Sequence 1568 BP; 258 A; 525 C; 490 G; 295 T; 0 other;  
Query Match 22.2%; Score 193.6; DB 22; Length 1568;  
Best Local Similarity 78.4%; Pred. No. 1.2e-46;  
Matches 232; Conservative 0; Mismatches 64; Indels 0; Gaps 0;  
QY 10 CAGCGGTTTGGTTAAGCCCATTTGATGGCAATCGTGTGACCTGGTTGAACCCGAAT 69  
DB 1177 CGGACGGTCTGGTTCAGGCCCATGCTCATGGCCATTTGTGACCTGGCTCAATCCCAAT 1236  
QY 70 GCCTATTTGGACCGTTTGTGTTTATCGCGCGCTGGCGCAATACGGCGACACCGGA 129  
DB 1237 GCCTACTGATGCCCTTCGCTTCATCGGCGGTGTGGAGCCAGTACGGGGAGACCGGT 1296  
QY 130 CGGTGATTTTCGCGCTGGCGGTTTCGCGGCAAGCTGATCTGGTTCCGCTGGTGGGT 189  
DB 1297 CGGTGATCTTCGCTGGGTCCTTCGCGCCAGCTGTCTGTTCCTCACTGGTCGGT 1356  
QY 190 TTGGCGCAGCAGCATTTGACGCCCGCTGTCCAGCCCAAGGTGTGGCGCTGGATCAAC 249

DB 1357 TACGGCGCGCCGCACTGTTCGGCTCCCTCTCCAGCCCGCGGTCTGGCGCTGGATCAAC 1416  
QY 250 GTCGTGTGGCACTTGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGGTTA 305  
DB 1417 ATAGGTGTGGCCGCTGTGCTCACCGGATTTGGCCGTGAAGCTGATCTGATGGGTTA 1472  
RESULT 9  
AAQ55755  
ID AAQ55755 standard; DNA; 5541 BP.  
XX AC AAQ55755;  
XX DT 25-SEP-1995 (first entry)  
XX DE Escherichia coli genomic probe EC-625.  
XX KW Probe; S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;  
XX KW K.pneumoniae; E.cloacae; clinical sample; ds.  
XX OS Escherichia coli.  
XX PN WO9401583-A.  
XX PD 20-JAN-1994.  
XX PF 07-JUL-1993; 93WO-JP00936.  
XX PR 07-JUL-1992; 92JP-0179719.  
XX PA (FUSO ) FUSO PHARM IND LTD.  
XX PI (OHNO ) OHNO T.  
XX PI Eda S, Matsuhisa A, Ohno T, Uehara H;  
XX WPI; 1994-035086/04.  
XX PT Probe for identifying bacteria causing infectious disease -  
PT consists of a DNA fragment obtained by HindIII cleavage of the  
PT pathogenic bacterial genomic DNA  
XX Claim 7; Page 70-73; 100pp; Japanese.  
XX The nucleotide sequence of a 5541 bp probe obtained by digestion of  
CC Escherichia coli genomic DNA with the restriction enzyme HindIII.  
CC The probes (AAQ55133-48 and AAQ55752-58) represent other probes derived,  
CC by HindIII digestion of the genomes of Staphylococcus aureus,  
CC S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,  
CC Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used  
CC to detect their respective microorganisms in clinical samples.  
XX SQ Sequence 5541 BP; 1331 A; 1353 C; 1417 G; 1440 T; 0 other;  
Query Match 8.2%; Score 71.6; DB 15; Length 5541;  
Best Local Similarity 49.4%; Pred. No. 1.8e-10;  
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;  
QY 203 CATTTGTCAGCCCGCTGTCCAGCCCAAGGTGTGGCGTGGATCAACGTCGTCGGCAG 262  
DB 2687 CATTATCCATCTGACGTTTGTGCTCTCTGCAATTTGTGATGGGCTATCTTGACCGACTGAC 2746  
QY 263 TTGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGGTTAGTTTCGGGGTTTGG 322  
DB\* 2747 TCGTCAATATCACTGATCTTATGCGGGCGCGGTTCTCGCGCCCTTATTAACAGGTCATT 2806  
QY 323 AATCGGTGGCTTTCGCCCAAAATGTTGATCGCGCGCTCGTGGGAAATCTCATCGATCGCT 382  
DB 2807 TATCGGAAGACCCCTGGCCACAGATTCAGCTGCCATCGCGGATGCTGATCAATCTGG 2866  
QY 383 CCAACTCGCGCTGAGAAAATCTCAAGTTTGTAGTGAATCAAGGCTGTGTCCAGCTGCT 442  
DB 2867 CCAGCTCTCTCGTGCTGCTAAATGTGAGATTATTTCAGCGCCTGCACGTTCTCTCTCAAGTTGTC 2926



OS Propionibacterium acnes.  
 PN WC200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 PR 02-JUN-2000; 2000US-208841P.  
 PR 07-JUL-2000; 2000US-215747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 XX  
 PS Claim 1; SEQ ID No 47; 1069pp; English.  
 XX  
 CC Sequences AAS59506-AAS59804 represent DNA molecules encoding  
 CC Propionibacterium acnes immunogenic polypeptides. The proteins and their  
 CC associated DNA sequences are used in the treatment, prevention and  
 CC diagnosis of medical conditions caused by P. acnes. The disorders include  
 CC SAPHO syndrome (synovitis, acne, pustulosis, hyperostosis and  
 CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved  
 CC in infections of bone, joints and the central nervous system, however it  
 CC is particularly involved in the inflammatory lesions associated with acne  
 CC vulgaris. A method for detecting the presence or absence of P. acnes in a  
 CC patient comprises contacting a sample with a binding agent that binds to  
 CC the proteins of the invention and determining the amount of bound protein  
 CC in the sample. The polypeptides may be used as antigens in the production  
 CC of antibodies specific for P. acnes proteins. These antibodies can be  
 CC used to downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the  
 CC polypeptides shown in AAUS1663-AAUS1893 and AAU67535.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 23128 BP; 4349 A; 6746 C; 7113 G; 4908 T; 12 other;  
 Query Match 6.6%; Score 57.6; DB 23; Length 23128;  
 Best Local Similarity 51.6%; Pred. No. 4.5e-06;  
 Matches 194; Conservative 1; Mismatches 160; Indels 21; Gaps 2;  
 QY 500 GCTCGGAGACCCATGCAAGCGCCATCTGCGCAAGTACTGCCCGTCTCGGCGGA 559  
 Db 7924 GCTCAGCAGGATGTCAGGAGGCGCATCTGGGCAACGCTGTGTCTGCTTTGCAA 7865  
 QY 560 TGTCATTGAGCTTGGGACCATATCAATATGTTTCAGTTCAACATGCCCTCAGACGG 619  
 Db 7864 CCTCGTTGAGGCGAGCAATTTGACGCGGGGTGACGTGCTGATCGGTCGACGCG 7805  
 QY 620 A---CTTACCTGCTGGCGCGGAACCCCTCTGGAATTCATCGAGATATTGTCGCTGA 676  
 Db 7804 AGACCGCTCCATGCCAGCAGCGACTGGGCGGGAACACCCCTCAGTATTGTCGGTGA 7745  
 QY 677 GCAGGCCCTGGCAAGTGGTGAAGAACAATGACGCCAAGACCACTTTGCGCACTCACT 736  
 Db 7744 GGATTCTCGTGAAGCAACGAGAGAAACAGATGATGCCATCCCTCTCTGCTCAGGTAT 7685  
 QY 737 GCACAAAGTTCTACCGTCTATCGCCCGGTTCTCCACCCCAAGATTAATGATGAATAGC 796  
 Db 7684 CGA-----CGAGACCATCTCTCAATCCATCGATCAACATGTTGTAAC 7643

QY 797 TTGGCTGATGAATCAGAAGCGGCGAGCCCTCTCCGCACTCAGCGCGCTCGCTG 856  
 Db 7642 GCGGCTGGTGGATGAGTACGGGGTGCCAGGTACGCGCAATGGTGGCAGACGTTCTG 7583  
 QY 857 TGAGCTCTGGACCGTA 872  
 Db 7582 TCTCCTTGGGATCGTA 7567  
 RESULT 12  
 AAT33536/c  
 ID AAT33536 standard; DNA; 15239 BP.  
 XX  
 AC AAT33536;  
 XX  
 DT 15-FEB-1998 (first entry)  
 XX  
 DE BCG deletion region 2 and flanking sequences.  
 XX  
 KW BCG delta 2; virulence; avirulence; attenuation; gene deletion;  
 KW mycobacteria; vaccine; infection; marker; ss.  
 XX  
 OS Mycobacterium bovis strain BCG.  
 XX  
 FH Key Location/Qualifiers  
 FT misc\_feature 3382..14071  
 FT /tag= a  
 FT /note= "BCG delta 1 deletion region"  
 PN W09625519-A1.  
 XX  
 PD 22-AUG-1996.  
 XX  
 PF 15-FEB-1996; 96WO-US01938.  
 XX  
 PR 17-FEB-1995; 95US-0390878.  
 XX  
 PA (PATH-) PATHOGENESIS CORP.  
 XX  
 PI Mahairas GG, Stover CK;  
 XX  
 DR WPI; 1996-393419/39.  
 XX  
 PT Detecting markers for avirulence in Mycobacterium - used in  
 PT production of vaccines against bacterial infection, and to detect  
 PT bacterial infection  
 XX  
 PS Example 1; Fig 2; 66pp; English.  
 XX  
 CC This DNA sequence comprises Mycobacterium bovis BCG deletion  
 CC sequence BCGdelta2. A specific genetic deletion of this region  
 CC results in an avirulence phenotype of the mycobacterium. 2 Other  
 CC deletion regions (see AAT33535 and AAT33537) have also been detected.  
 CC Identification involved screening a BCG cosmid library with a  
 CC radiolabeled probe obtained following DNA subtraction between  
 CC virulent Mycobacterium tuberculosis H37Rv and avirulent BCG.  
 CC The deletions provide useful markers for the identification of an  
 CC avirulent, or a virulent, mycobacterial phenotype. Determination  
 CC of avirulence requires the detection of the presence or absence of  
 CC the deletion; the deletions are detected either by detecting the  
 CC presence or absence of deletion junctions (see AAT33538-46), or by  
 CC detecting the presence or absence of the sequences contained within  
 CC the deletion. Deletion polypeptides are used as components of  
 CC immunological assays and in vaccines.  
 XX  
 SQ Sequence 15239 BP; 2878 A; 4791 C; 4729 G; 2841 T; 0 other;  
 Query Match 6.5%; Score 57; DB 17; Length 15239;  
 Best Local Similarity 53.1%; Pred. No. 5.6e-06;  
 Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;  
 QY 26 AGCCCATGTTGATGGCAATCGTGTGACCTGTTGAACCCGAATCGTATTGGACGCT 85

Db 4792 AGGTCCTGGTGAACCTGTGGCGCATTCACGTTCTCAACCCACACAGCTACTCTCGACACCG 4733  
 QY 86 TTGTGTTTATCGCGCGGTGCGCGCGCAATACGGCGACACGGAGCGTGTGATTTTCGCCG 145  
 Db 4732 TCGGTGTTGCTAGGCGCGCTGGCCAAACGAGCACACGGA---CCAGCGCTGGCTGTTCGCCG 4676  
 QY 146 CTGCGCGGTTTCGCGCGCAAGCTGTATCTGTTTCCCGCTGGTGGGTTTCGGCGCAGCAGCAT 205  
 Db 4675 TCGCGCGGTTTCACAGCAGTGGGTATGTTTCGCCACCTTCGGGTTTCGGAGCGCGCGGT 4616  
 QY 206 TGTCAACCGCGCTGTTCAGACCCCAAGGTGTGGCGCTGGATCAACGTCTGTGTCGCAAGTTG 265  
 Db 4615 TCGCGCGGCTGTTTCAACCAACCCCGGCTGTGGAGAACTCTCGACGGCTGTATCGCGGTCA 4556  
 QY 266 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 298  
 Db 4555 TGATGGTTGGCTGGGAATCTCGCTGACCCGTGA 4523

## RESULT 13

AAI99683  
 ID AAI99683 standard; DNA; 4403765 BP.

XX AC AAI99683;

DT 15-JAN-2002 (first entry)

XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX Mycobacterium tuberculosis.

XX US6294328-B1.

XX 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

XX 24-JUN-1998; 98US-0103840.

XX (GENO-) INST GENOMIC RES.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ

XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
 Query Match 6.5%; Score 57; DB 22; Length 4403765;  
 Best Local Similarity 53.1%; Pred. No. 6.3e-05;

Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;  
 QY 26 AGCCCAATGTTGATGGCAATCGCTGACCTGTTGAACCCGAATCGGTATTTGGAGCGGT 85  
 Db 2227633 AGGTCTCTGTGACCTGTGGCGCATTCACGTTCTCAACCCACACAGCTCTACCTCGACACCG 2227692  
 QY 86 TTGTGTTTATCGCGCGGTGCGCGCGCAATACGGCGACACCGGAGCGTGTGATTTTCGCCG 145  
 Db 2227693 TCGTGTCTAGGCGCGCTGGCCAAACGAGCACAGGA---CCAGCGCTGGCTGTTCGCCG 2227749  
 QY 146 CTGCGCGGTTTCGCGCGCAAGCTGTATCTGTTTCCCGCTGGTGGGTTTCGGCGCAGCAGCAT 205  
 Db 2227750 TCGCGCGGTTTCACAGCAGTGGGTATGTTTCGCCACCTTCGGGTTTCGGAGCGCGCGGT 2227809  
 QY 206 TGTCAACCGCGCTGTTCAGACCCCAAGGTGTGGCGCTGGATCAACGTCTGTGCGCAAGTTG 265  
 Db 2227810 TCGCGCGGCTGTTCACCAACCCCGGCTGTGGAGAACTCTCGACGGCTGTATCGCGGTCA 2227869  
 QY 266 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 298  
 Db 2227870 TGATGGTTGGCTGGGAATCTCGCTGACCCGTGA 2227902

## RESULT 14

AAI99682  
 ID AAI99682 standard; DNA; 4411529 BP.

XX AC AAI99682;

DT 15-JAN-2002 (first entry)

XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

XX Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.

XX Mycobacterium tuberculosis.

XX US6294328-B1.

XX 25-SEP-2001.

XX 24-JUN-1998; 98US-0103840.

XX 24-JUN-1998; 98US-0103840.

XX (GENO-) INST GENOMIC RES.

XX Fleischmann RD, White OR, Fraser CM, Venter JC;

XX WPI; 2001-647261/74.

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 PT 1551 and H37Rv differ

XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

XX

PT Increased production of L-amino acids by an *Escherichia* bacterium  
PT comprises increasing the expression amount of an L-amino acid excretion  
PT protein -  
XX  
XX Disclosure: Page 24; 29pp; English.  
PS  
PS  
XX The present sequence is the yggA gene (an excretion protein gene) of  
XX *Escherichia coli*. The amino acid excretion protein produced from this  
CC gene is involved in the production of amino acids, and an increase in its  
CC  
CC

Search completed: April 27, 2003, 00:06:32  
Job time : 5330.11 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 00:51:48 ; Search time 98.593 Seconds  
(without alignments)  
9634.930 Million cell updates/sec

Title: US-09-105-117K-1\_COPY\_1421\_2293

Perfect score: 873

Sequence: 1 gtcgataagcagcggttg.....ctgtgagctgacgcgtag 873

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	873	100.0	3309400	9	US-09-738-626-1
C 2	568	65.1	627	9	US-09-738-626-3454
C 3	326	37.3	822	9	US-09-746-660A-51
C 4	303	34.7	708	9	US-09-738-626-3455
C 5	65.4	7.5	988	10	US-09-974-300-77
C 6	57	6.5	597	9	US-09-894-844-11
C 7	50.2	5.8	1277	9	US-09-971-536-15
C 8	40	4.6	399	10	US-09-861-893-29
C 9	40	4.6	431	10	US-09-954-456-1034
C 10	38.2	4.4	948	9	US-09-738-626-1672
C 11	36.4	4.2	499	9	US-10-184-644-592
C 12	36.4	4.2	499	9	US-10-184-634-592
C 13	36	4.1	1068	9	US-09-738-626-439
C 14	36	4.1	3309400	9	US-09-738-626-1
C 15	35.4	4.1	996	10	US-09-815-242-7775
C 16	34.8	4.0	542	9	US-10-184-644-398
C 17	34.8	4.0	542	9	US-10-123-155-188
C 18	34.8	4.0	542	9	US-10-184-634-398
C 19	34.8	4.0	3017	10	US-09-818-143-8

20	34.8	4.0	7158	10	US-09-974-300-2171	Sequence 2171, Ap
21	33.8	3.9	887	10	US-09-864-761-21484	Sequence 21484, A
22	33.8	3.9	1981	10	US-09-864-761-4745	Sequence 4745, Ap
23	33.8	3.9	2930	10	US-09-745-763-198	Sequence 198, App
24	33.6	3.8	1075	10	US-09-864-761-19241	Sequence 19241, A
25	33.6	3.8	1403	10	US-09-864-761-2513	Sequence 2513, Ap
26	33.6	3.8	1446	9	US-09-712-363-57	Sequence 57, Appl
27	33.2	3.8	1635	10	US-09-864-761-20241	Sequence 20241, A
28	33.2	3.8	1973	10	US-09-864-761-3471	Sequence 3471, Ap
29	33	3.8	9210	9	US-09-712-363-100	Sequence 100, App
30	32.8	3.8	2048	10	US-09-925-300-760	Sequence 760, App
31	32.4	3.7	4689	9	US-09-860-846-34	Sequence 34, Appl
32	32.4	3.7	4689	9	US-09-988-3848-34	Sequence 34, Appl
33	32.4	3.7	4689	10	US-09-861-289-34	Sequence 34, Appl
34	32.4	3.7	36778	9	US-09-860-846-5	Sequence 5, Appl
35	32.4	3.7	36778	10	US-09-861-289-5	Sequence 5, Appl
36	32.4	3.7	37948	9	US-09-988-3848-5	Sequence 5, Appl
37	32.2	3.7	484	9	US-09-918-995-3113	Sequence 3113, Ap
38	32.2	3.7	497	9	US-09-918-995-27219	Sequence 27219, A
39	32.2	3.7	653	9	US-10-184-644-402	Sequence 402, App
40	32.2	3.7	653	9	US-10-184-634-402	Sequence 402, App
41	32.2	3.7	1404	9	US-09-976-740-10	Sequence 10, Appl
42	32.2	3.7	1404	10	US-09-962-055-10	Sequence 10, Appl
43	32.2	3.7	1404	12	US-10-023-529-10	Sequence 10, Appl
44	32.2	3.7	1404	12	US-10-023-523-10	Sequence 10, Appl
45	32	3.7	360	9	US-10-184-644-28	Sequence 28, Appl

#### ALIGNMENTS

##### RESULT 1

US-09-738-626-1/c  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 100.0%; Score 873; DB 9; Length 3309400;

Best Local Similarity 100.0%; Pred. No. 7.6e-281;

Matches 873; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGATAAGCAGCGGTTTGGTAAAGCCCATGTTGATGGCAATCGTGTACCTGGTTG 60

Db\_1328548 GTCGATAAGCAGCGGTTTGGTAAAGCCCATGTTGATGGCAATCGTGTACCTGGTTG 1328489

QY 61 AACCCGAATGCGTATTGTGACCGGTTTGTATTATCGGCGCGCTCGCGCGCAATACGCG 120



[illegible]

```

RESULT 2
US-09-738-626-3454/c
/ Sequence 3454, Application US/09738626
/ Publication No. US20020197605A1
/ GENERAL INFORMATION:
/ APPLICANT: NAKAGAWA, SATOSHI
/ APPLICANT: MIZOGUCHI, HIROSHI
/ APPLICANT: ANDO, SEIKO
/ APPLICANT: HAYASHI, MIKIRO
/ APPLICANT: OCHITAI, KEIKO
/ APPLICANT: YOKOI, HARUHIKO
/ APPLICANT: TATEISHI, NAOKO
/ APPLICANT: SENOH, AKIHIRO
/ APPLICANT: IKEDA, MASATO
/ APPLICANT: OZAKI, AKIO
/ TITLE OF INVENTION: NOVEL POLYNUCLEOT
/ FILE REFERENCE: 249-125

```

```

; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIORITY FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3454
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3454

Query Match          65.1%; Score 568; DB 9; Length 627;
Best Local Similarity 100.0%; Pred. No. 3e-180;
Matches 568; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 306 GTTTTTCGCGGTTTGAATCGGTGGCTTCGCGCCAAATGTTGATGTCGCGCGCTCGTGGA 365
DB 627 GTTTTTCGCGGTTTGAATCGGTGGCTTCGCGCCAAATGTTGATGTCGCGCGCTCGTGGA 568

QY 366 AATCTCATCGATCGCTCCAACTCGGCGTCGAGAAAATCTCAAAGTTGTTGAGTGAATCAAG 425
DB 567 AATCTCATCGATCGCTCCAACTCGGCGTCGAGAAAATCTCAAAGTTGTTGAGTGAATCAAG 508

QY 426 GCTGTTGTCAGTGTCTCAATGACGAAGACCAATCAATGCACTGGTCACTGGTATCCGC 485
DB 507 GCTGTTGTCAGTGTCTCAATGACGAAGACCAATCAATGCACTGGTCACTGGTATCCGC 448

QY 486 GCGTACTCTCTTGTCTGCGGACGACCCATGCAAGCGCCATCTGCGCAAGTCACTGCC 545
DB 447 GCGTACTCTCTTGTCTGCGGACGACCCATGCAAGCGCCATCTGCGCAAGTCACTGCC 388

QY 546 GCGTTCCTGGCGCATGTTCATTGAGCTTTCGCGACCATATCAATATTGTTCACTGTTCAACAT 605
DB 387 GCGTTCCTGGCGCATGTTCATTGAGCTTTCGCGACCATATCAATATTGTTCACTGTTCAACAT 328

QY 606 GCGCTCAGACAGGACATTACCTGCTGGCTGGCGGGAACCTCTGGAATTCATCCATCGAGATA 665
DB 327 GCGCTCAGACAGGACATTACCTGCTGGCTGGCGGGAACCTCTGGAATTCATCCATCGAGATA 268

QY 666 TTTGTCGCTGACGAGCGCTCCGCAAGTGTGTGAGAAAGCAATGACGCCAGACATTGTT 725
DB 267 TTTGTCGCTGACGAGCGCTCCGCAAGTGTGTGAGAAAGCAATGACGCCAGACATTGTT 208

QY 726 GGCAGCTGACTGCAACAAAGTTCTCACCGTCATCGCCGGTTCCTCCACCCAAAGATTAAAT 785
DB 207 GGCAGCTGACTGCAACAAAGTTCTCACCGTCATCGCCGGTTCCTCCACCCAAAGATTAAAT 148

QY * 786 GATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGACCCCTCTCTCGGCCATGAACCTCAGC 845
DB 147 GATGGAATAGCTTGGCTGATGAATCAGAAGCGGCGACCCCTCTCTCGGCCATGAACCTCAGC 88

QY 846 GCGCTCCGCTGTGAGCTCTGAGCCGTAG 873
DB 87 GCGCTCCGCTGTGAGCTCTGAGCCGTAG 60

RESULT 3
US-09-746-660A-51
; Sequence 51, Application US/09746660A
; Publication No. US20030049804A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zeider, Oskar
; APPLICANT: Haberhauer, Gregor
; APPLICANT: Kim, Jun-Won
; APPLICANT: Lee, Heung-Schick

```

APPLICANT: Hwang, Byung-Joon  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
FILE REFERENCE: BGI-121C2  
CURRENT APPLICATION NUMBER: US/09/746,660A  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/606740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 09/603124  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/142101  
PRIOR FILING DATE: 1999-07-02  
PRIOR APPLICATION NUMBER: 60/148613  
PRIOR FILING DATE: 1999-08-12  
PRIOR APPLICATION NUMBER: 60/187970  
PRIOR FILING DATE: 2000-03-09  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
NUMBER OF SEQ ID NOS: 125  
SOFTWARE: PatentIn Vers. 2.0  
SEQ ID NO 51  
LENGTH: 822  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (101)..(799)  
OTHER INFORMATION: RXA01394  
US-09-746-660A-51

Query Match 37.3%; Score 326; DB 9; Length 822;  
Best Local Similarity 100.0%; Pred. No. 5.4e-99;  
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGATTAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 60  
Db 497 GTCGATTAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 556  
QY 61 AACCCGAATCGTATTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGGC 120  
Db 557 AACCCGAATCGTATTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGGC 616  
QY 121 GACACCGAGCGTGGATTTTCGCGCGTGGCGGTTTCGCGGCAAGCTGATCTGGTTCCCG 180  
Db 617 GACACCGAGCGTGGATTTTCGCGCGTGGCGGTTTCGCGGCAAGCTGATCTGGTTCCCG 676  
QY 181 CTGTGGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCAGCCCAAGGTTGGCGC 240  
Db 677 CTGTGGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCAGCCCAAGGTTGGCGC 736  
QY 241 TGGATCAACGTCGTCTGTCGTCAGTTCGATGACCGCATGGCCATCAAACTGATGTTGATG 300  
Db 737 TGGATCAACGTCGTCTGTCGTCAGTTCGATGACCGCATGGCCATCAAACTGATGTTGATG 796  
QY 301 GGTTAGTTTCGCGCGGTTTGGGAATC 326  
Db 797 GGTTAGTTTCGCGCGGTTTGGGAATC 822

RESULT 4  
US-09-738-626-3455  
Sequence 3455, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIALI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAKO

APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 3455  
LENGTH: 708  
TYPE: DNA  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3455

Query Match 34.7%; Score 303; DB 9; Length 708;  
Best Local Similarity 100.0%; Pred. No. 2.6e-91;  
Matches 303; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGATTAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 60  
Db 406 GTCGATTAAGCAGCGGTTTGGTAAAGCCCATGTTGATGCGCAATCGTCTGACCTGGTTG 465  
QY 61 AACCCGAATCGTATTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGGC 120  
Db 466 AACCCGAATCGTATTGGACGCGTTTGTGTTATCGCGCGCGTGGCGGCGCAATACGGC 525  
QY 121 GACACCGAGCGTGGATTTTCGCGCGTGGCGGTTTCGCGGCAAGCTGATCTGGTTCCCG 180  
Db 526 GACACCGAGCGTGGATTTTCGCGCGTGGCGGTTTCGCGGCAAGCTGATCTGGTTCCCG 585  
QY 181 CTGTGGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCAGCCCAAGGTTGGCGC 240  
Db 586 CTGTGGGTTTCGCGCGCAGCAGCATTTGTCACGCCGCTGTCAGCCCAAGGTTGGCGC 645  
QY 241 TGGATCAACGTCGTCTGTCGTCAGTTCGATGACCGCATGGCCATCAAACTGATGTTGATG 300  
Db 646 TGGATCAACGTCGTCTGTCGTCAGTTCGATGACCGCATGGCCATCAAACTGATGTTGATG 705  
QY 301 GGT 303  
Db 706 GGT 708

RESULT 5  
US-09-974-300-77/c  
Sequence 77, Application US/09974300  
Patent No. US20020146721A1  
GENERAL INFORMATION:  
APPLICANT: Berka, Randy M.  
APPLICANT: Clausen, Ib Groth  
TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
FILE REFERENCE: 10085.500-US  
CURRENT APPLICATION NUMBER: US/09/974,300  
CURRENT FILING DATE: 2001-10-05  
PRIOR APPLICATION NUMBER: 09/680,598  
PRIOR FILING DATE: 2000-10-06  
PRIOR APPLICATION NUMBER: 60/279,526  
PRIOR FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 8481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 77  
LENGTH: 988  
TYPE: DNA  
ORGANISM: Bacillus licheniformis  
US-09-974-300-77

Query Match 7.5%; Score 65.4; DB 10; Length 988;  
Best Local Similarity 55.3%; Pred. No. 1.6e-11;  
Matches 204; Conservative 0; Mismatches 141; Indels 24; Gaps 3;

QY 511 ACCCATGCAAGCGCATCTGGCAAGTGACTGCCCGGTTCTCTGGGCGATGTCATTGAGC 570  
DB 857 AGCAGGCTATCGCATTTGGGGAAGAGACTGGCCGGTTGGGAGCAGCATTTGATTGAGC 798  
QY 571 TTGGGACCATATCAATATTGTTTCCAGCTTCAACATGCCCTCAGACAGGACTTAC----- 625  
DB 797 TTGGGACCTTGGCGAGCAGCTCTTCTGTGATGCCCTCGCGTCAAAAAGCGACTGTTTC 738  
QY 626 -CCTGGCTGGGGGGGAACCTCTGGATTCATCGAGATATTGTCGGTGAGAGGCC 684  
DB 737 GCATCAGCGGCTCTTGACCCGGCGGGAATCCCTGTCAATATTCTTGTGTTTAAAGCCCT 678  
QY 685 TGCCCAAGTGTGAGAAAGCAATCAAGCCCAAGCAATGTTGGCAGCTGACTGCAACAAG 744  
DB 677 TGGCAAGAGGTGTAAGCGATGCCCGGCTCCTTTT-----CTGAGGACT 627  
QY 745 TTCTACCGTCATCGCCCGGTTCTTCCACCAACGATTAATGATGGAATAGCTTGGCTGA 804  
DB 626 TGCAGCAGGCTTTTTCGGGTTCT-----CTGTGGAACATTGAATATTGGGCTGA 576  
QY 805 TGATCAGAGCGGCGAGCCCTCTCGCCATGAACTCAGCGGCTCGGCTGTGAGCTCT 864  
DB 575 TGGATGAGGAGCGGAATCCCGTATTCTTGCAAAATATCAGCGGCTTTTTCGGTCTGTTCA 516  
QY 865 GGACCGTAG 873  
DB 515 GCATGCTAG 507

RESULT 6  
US-09-894-844-11  
; Sequence 11, Application US/09894844  
; Patent No. US20020176873A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of  
; TITLE OF INVENTION: the M. Tuberculosis Complex  
; FILE REFERENCE: STAN102CON  
; CURRENT APPLICATION NUMBER: US/09/894,844  
; CURRENT FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Mycobacteria tuberculosis  
; US-09-894-844-11

Query Match 6.5%; Score 57; DB 9; Length 597;  
Best Local Similarity 53.1%; Pred. No. 8.4e-09;  
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 26 AGCCCATGTTGATGGCAATCGTGTGACCTGGTTGAACCCGGAATTCGTAATTGGACCGCT 85  
DB 326 AGGTCTGTTGACCTGTGCGGCAATTCAGTTCTCAACCCACAGCTTACTCTGACACCG 385  
QY 86 TTGTGTTTATCGGCGGCTCGCGCGGCAATACAGCGGACACCGGACGTTGATTTTCGCG 145  
DB 386 TCGGTTGTTGCTAGGCGCGCTGGCCCAACGAGCACAGCGA---CCAGCGCTGGCTGTTTCGCC 442  
QY 146 CTGGCGGCTTTCGCGCAAGCCTGATCTGTTCCCGCTGGTGGTTCGCGCGCAGCAGCAT 205

DB 443 TGGCGCGGCTCACAGCAGTGGGTATGTTGCCACCCCTCGGGTTCGAGCGCGCGGT 502  
QY 206 TGTACAGCCCGCTGTCTCAGCCCAAGGTGTGGCGTGGATCAACGTCGTCGTGGCAGTTG 265  
DB 503 TGGCGGGGCTGTTTCAACCAACCCCGGCTCGTGAGAAATCTCGACGCGCTGATCGGGTCA 562  
QY 266 TGATGACCGCATTTGGCCATCAAACTGATTTGA 298  
DB 563 TGATGTTGCGCTGGGAATCTCGCTGACCGTGA 595

RESULT 7  
US-09-971-536-15/c  
; Sequence 15, Application US/09971536  
; Patent No. US20020159976A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka  
; APPLICANT: Bloksberg, Leonard  
; APPLICANT: Lubbers, Mark  
; APPLICANT: Dekker, James  
; APPLICANT: Christensen, Anna  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul  
; APPLICANT: Reid, Julian  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods  
; TITLE OF INVENTION: Using Them  
; FILE REFERENCE: 1043c2  
; CURRENT APPLICATION NUMBER: US/09/971,536  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 1277  
; TYPE: DNA  
; ORGANISM: Lactobacillus rhamnosus  
; US-09-971-536-15

Query Match 5.8%; Score 50.2; DB 9; Length 1277;  
Best Local Similarity 49.3%; Pred. No. 2.3e-06;  
Matches 176; Conservative 0; Mismatches 163; Indels 18; Gaps 1;

QY 365 AATCTCATGATCGCTCCAACTCGGCGTCAGAAAACCTCCAAAGTTGTTGAGTGAATCAA 424  
DB 1156 AATATCATCAATTTGTTGAATCTCTTACAGCAGTAAAGGTGAGTGTTCGTTGCTTAA 1097  
QY 425 GGCTGTTGTCAGCTCTCACTGACGAGCAGCACCATTCAATGACCTGCTCAAGTATCCG 484  
DB 1096 GGTATCTGGAAGGTGTTCACTGAGTCTGTCCTCAATGATCACTGTGACAA----- 1043  
QY 485 CGCCGTACTCTCTCTGCGCGCAGCAGCAGCCATCGAAGCGCCATCTCGCAAGTGAATGCC 544  
DB 1042 -----CCGGATCCCGTAACAACACCGCCAAAGGCCATTGACTCAGGTTTGGT 995  
QY 545 CGGTTCTCGGGGATGTCATTGAGCTTGGGACCATATCAATATTGTTCAAGTCAACA 604  
DB 994 CAGCATATGCGCAATTTTCAATTAAGCGCATTTAGTTGCTTAAACACAGCGCTCTTTGCCCT 935  
QY 605 TGCCCTCAGACAGGAGTACCTCGCTGCGCGGAACCTCTGGAATTTCCATCGAGAT 664  
DB 934 TAGCAAAAGTGGCTTGTGTTGGATGGAATTTTGAAGTATCCGGAATTTCCCTTTAGT 875  
QY 665 ATTTGTCGTCAGCAGCGCTCGCGCAAGTGGTGAGAAAGCAATGACGCCAAGACCAT 721  
DB 874 AGCGATCTGATACAAAGCCTTCTGTATAACGTCGATGCAATCAACCCGCGACCAT 818

```

; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1034
; LENGTH: 431
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1034

Query Match          4.6%; Score 40; DB 10; Length 431;
Best Local Similarity 53.1%; Pred. No. 0.0036;
Matches 85; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 211 CGCCCGCTGTCCAGCCCCAAGGTGGCGCTGGATCAACGTCGTCGCGCAGTTGTGATG 270
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 CGCCCGCTGGTCAACCGCAGCGTCCCGAGTCGCTCTCGGGCGTGTGTCTTGATGAAG 61
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 271 ACCGATTGGCCATCAAACTGATTTGATGGTTAGTTTTTCGCGGGTTTTTGGAAATCGGTG 330
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 TAGTGCCTGTCTTGGCCCTCGATCGTGAAGTCAGGTTCTCCAGGTACAGCGCGTTGTTG 121
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 331 GCCTTCGCCCAATGTTGATGCCGCGCTCGTGGGAAATCT 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 AGCACGGCGCCACCTTGATCAGTCCTCGTTGGCGATGT 161
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-738-626-1672/c
; Sequence 1672, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1672
; LENGTH: 948
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1672

Query Match          4.4%; Score 38.2; DB 9; Length 948;
Best Local Similarity 63.7%; Pred. No. 0.021;
Matches 58; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 634 GCGCGGAAACCTCTGGAATTCATCGAGATATTTGTCGTGAGCAGCGCCTGCGCAAGT 693
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 689 GCTCTGAATCATGGGAAATTCGAGCGGTATTTAGCAGTCAGCACGCGCTTGCCCGCAGC 630
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 694 GGTGAGAAAGCAATGACGCCAAGACCAATTGT 724
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 629 GGAGCGCCAGCAAGAATCCGACACCTTAGT 599
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-439

Query Match 4.1%; Score 36; DB 9; Length 1068;  
Best Local Similarity 53.6%; Pred. No. 0.12; Mismatches 0; Gaps 0;  
Matches 75; Conservative 0; Indels 65; Indels 0; Gaps 0;

QY 146 CTGGCGGCTTCGGCGCAAGCTGATCTGGTTCGGCTGGTGGTTCGGCGCAGCAGCAT 205  
DB 542 CCGCGCGCTGACCTGCTTCTTATCTACATCGCGCGCAGAGCTCCCGCGAAGAAATTA 601  
QY 206 TGTACAGCCCGCTGTCAGCCCAAGTGTGGCGCTGGATCAAGTGTGTCGGCAGTTG 265  
DB 602 TTTTCTGGCAGATGGGTTCCTCAACGGCTCGCAATGGCGCGCAGCTCAACGTGGTGTG 661  
QY 266 TGATGACCGCATTTGGCCATC 285  
DB 662 TGATCGTGAATCGGCTC 681

## RESULT 14

US-09-738-626-1  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 4.1%; Score 36; DB 9; Length 3309400;  
Best Local Similarity 53.6%; Pred. No. 5.1;  
Matches 75; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 146 CTGGCGGCTTCGGCGCAAGCTGATCTGGTTCGGCTGGTGGTTCGGCGCAGCAGCAT 205  
DB 414184 CCGCGCGCTGACCTGCTTCTTATCTACATCGCGCGCAGAGCTCCCGCGAAGAAATTA 414243  
QY 206 TGTACAGCCCGCTGTCAGCCCAAGTGTGGCGCTGGATCAAGTGTGTCGGCAGTTG 265  
DB 414244 TTTTCTGGCAGATGGTTCCTCAACGGCTCGCAATGGCGCGCAGCTCAACGTGGTGTG 414303  
QY 266 TGATGACCGCATTTGGCCATC 285  
DB 414304 TGATCGTGAATCGGCTC 414323

## RESULT 15

US-09-815-242-7775/c  
; Sequence 7775, Application US/09815242

; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE OF INVENTION: Prokaryotes  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7775  
; LENGTH: 996  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(996)  
US-09-815-242-7775

Query Match 4.1%; Score 35.4; DB 10; Length 996;  
Best Local Similarity 53.2%; Pred. No. 0.19;  
Matches 75; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 470 TGGTCAAGTATCCGCGCGTACTCTCTTGTCTGGCAGCAGCCCATGCAAGCGCATCT 529  
DB 847 TGGTCCCGGAATCGGATCAGGTATCGCCCTGCGCAGCAGCCAGGCGCAGTT 788  
QY 530 GCGCAAGTGACTGCCCGGCTTCTCTGGCGATGTCTATTGAGCTTGGGACCATATCAATAT 589  
DB 787 GCGAGGGTTTACGGCCCTTGGCTGGCGCAGTTGCCGACCTTGTCCACAGCTTCAGGT 728  
QY 590 TGTTCAGTTCAACATGCCCT 610  
DB 727 TCTTCGCGAAGTTCTCGCCCT 707

Search completed: April 27, 2003, 05:34:59  
Job time : 3284.59 secs



Result No.	Score	Query			ID	Description	
		Match	Length	DB			
1	71.6	8.2	5541	1	US-08-920-812-20	Sequence 20, Appl	
2	71.6	8.2	5541	1	US-08-920-827-20	Sequence 20, Appl	
3	71.6	8.2	5541	1	US-08-921-177-20	Sequence 20, Appl	
4	71.6	8.2	5541	1	US-08-362-577C-20	Sequence 20, Appl	
5	71.6	8.2	5541	2	US-08-920-828-20	Sequence 20, Appl	
C	57	6.5	15239	1	US-08-390-878-17	Sequence 17, Appl	
	7	57	6.5	4403765	4	US-09-103-840A-2	Sequence 2, Appl
8	57	6.5	4411529	4	US-09-103-840A-1	Sequence 1, Appl	
C	9	39.6	4.5	1935	2	US-08-492-027A-9	Sequence 9, Appl
C	10	37.6	4.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl
C	11	37.6	4.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl
C	12	34.4	3.9	2048	3	US-08-776-246-1	Sequence 1, Appl
C	13	34.4	3.9	2048	3	US-08-776-246-1	Sequence 1, Appl
C	14	33.6	3.8	1681	4	US-09-434-288-7	Sequence 7, Appl
C	15	33.6	3.8	1681	4	US-07-945-283-1	Sequence 1, Appl
C	16	33.2	3.8	20235	1	US-07-642-734C-3	Sequence 3, Appl
C	17	33.2	3.8	20235	1	US-08-439-009A-3	Sequence 3, Appl
C	18	33	3.8	707	4	US-08-998-416-940	Sequence 940, App
C	19	32.8	3.8	4508	5	UCT-US93-06251-34	Sequence 34, Appl
C	20	32.6	3.7	421	3	US-08-782-480-7	Sequence 7, Appl
C	21	32.6	3.7	421	4	US-08-954-211-7	Sequence 7, Appl
C	22	32.6	3.7	421	4	US-09-005-167A-7	Sequence 7, Appl
C	23	32.6	3.7	421	4	US-09-176-741B-7	Sequence 7, Appl
C	24	32.4	3.7	1680	4	US-09-603-185-1	Sequence 1, Appl
C	25	32.4	3.7	4689	4	US-09-103-537-34	Sequence 34, Appl
C	26	32.4	3.7	36778	4	US-09-105-537-5	Sequence 5, Appl
C	27	32.4	3.7	38506	3	US-09-320-878-19	Sequence 19, Appl



Query Match 8.2%; Score 71.6; DB 1; Length 5541;  
 Best Local Similarity 49.4%; Pred. No. 1.7e-11; Indels 46; Gaps 4;  
 Matches 337; Conservative 0; Mismatches 299;

QY 203 CATTGTGACGCGCTGCTCCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTGTCGAC 262  
 DB 2687 CATTATCCATCTGACGTTTGTGCTCTCTGCAATTTGTGATGGCTATCTTGACCGACTGAC 2746

QY 263 TTGTGATGACCGCATGCGCATCAAACTGATGTTGATGTTGTTGTTTTCGGGGTTTGG 332  
 DB 2747 TCGTCATAATCACTGATCTTATGCGGCGGCTTCTCGCGCGCTTATTAACAGGTCATT 2806

QY 323 AATCGGTGGCTTCGCCCAATGTTGATGCGGGCTGTGGGAAATCTCATGATCGCCT 382  
 DB 2807 TATCGGAAGACGCGTCCACAGATTCAGCTCGCCATCGCGGATATGCTGATCAATCTCGG 2866

QY 383 CCAACTCGGCGTCAGAAACTTCCAAGTTGTTGATGTAATCAAGGCTGTGTCACAGCTGCT 442  
 DB 2867 CCAGCTCTCGGTCTAAATGTCAGATTAATTCAGCGCTGCACGTTCTCTCAAGTTGTC 2926

QY 443 CAACTGAGAAAGACCAATCAATGCACTGTGTCAGGTTATCGGCGGCTACTCTCTTGTCT 502  
 DB 2927 CGCGCGCTCGGCGGCTTATGCGGCGGCTTCTCGCGCGCTTATTAACAGGTCATT 2967

QY 503 CGCGCAGCACCATGCAAGCGCCATCTGCGCAAGTGTGCTGCGGCTTCTGCGCGATGT 562  
 DB 2968 CTTTCAGCAACAGCTTAAACGCCATTTGCGCCATTTGTCGCAATGTCGTCGTCGCAATTT 3027

QY 563 CATTGAGCTTCGCGACCATATCAATATTTGTTTCAAGTTTCAACATGCCCTCAGACAGGACT 622  
 DB 3028 CATTCAATAAGTGTAGGCTGTGAGTTGCTTTCGTAAGCAATTTTCGGCGTCAGACCA 3087

QY 623 TACCTGGCTGGC-----CGGGAAACCTCTCGGAATTCATCGAGATTTTGT 670  
 DB 3088 GAACCTTTATTTCCCTTTCACGATGTCATCGGAAATCTTGGGAAATCCGCTTGAGATTTTC 3147

QY 671 CCGTGACGAGCGCTGCGCAAGTGTGAGAAAGCAATGACGCAAGCAATTTGTCGAC 730  
 DB 3148 CGGTGAGCAATTCCTGAGCAGAGGATTAAGGCAATACAGCCCGCTTATTTTGA 3207

QY 731 CTGACTGCAACAGTTCTCACCGTCAATGCACTGGTCAAGGTTATCGGCGGCTACTCTCTTGTCT 790  
 DB 3313 TTTCGCGCTGCTGCGGAGTA 3334

## RESULT 2

US-08-920-827-20  
 ; Sequence 20, Application US/08920827  
 ; Patent No. 5770375  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/920,827  
 FILING DATE: 29-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/362,577  
 FILING DATE: 27-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rin-Laures, Li-Hsien  
 REGISTRATION NUMBER: 33,547  
 REFERENCE/DOCKET NUMBER: 19036/32420  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5541 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: Clinical Isolate EC-625  
 US-08-920-827-20

Query Match 8.2%; Score 71.6; DB 1; Length 5541;  
 Best Local Similarity 49.4%; Pred. No. 1.7e-11; Indels 46; Gaps 4;  
 Matches 337; Conservative 0; Mismatches 299;

QY 203 CATTGTGACGCGCTGCTCCAGCCCAAGGTGTGGCGCTGGATCAACGTCGTGTCGAC 262  
 DB 2687 CATTATCCATCTGACGTTTGTGCTCTCTGCAATTTGTGATGGCTATCTTGACCGACTGAC 2746

QY 263 TTGTGATGACCGCATGCGCATCAAACTGATGTTGATGTTGTTTTCGGGGTTTGG 332  
 DB 2747 TCGTCATAATCACTGATCTTATGCGGCGGCTTCTCGCGCGCTTATTAACAGGTCATT 2806

QY 323 AATCGGTGGCTTCGCCCAATGTTGATGCGGGCTGTGGGAAATCTCATGATCGCCT 382  
 DB 2807 TATCGGAAGACGCGTCCACAGATTCAGCTCGCCATCGCGGATATGCTGATCAATCTCGG 2866

QY 383 CCAACTCGGCGTCAGAAACTTCCAAGTTGTTGATGTAATCAAGGCTGTGTCACAGCTGCT 442  
 DB 2867 CCAGCTCTCGGTCTAAATGTCAGATTAATTCAGCGCTGCACGTTCTCTCAAGTTGTC 2926

QY 443 CAACTGAGAAAGACCAATCAATGCACTGTGTCAGGTTATCGGCGGCTACTCTCTTGTCT 502  
 DB 2927 CGCGCGCTCGGCGGCTTATGCGGCGGCTTCTCGCGCGCTTATTAACAGGTCATT 2967

QY 503 CGCGCAGCACCATGCAAGCGCCATCTGCGCAAGTGTGCTGCGGCTTCTGCGCGATGT 562  
 DB 2968 CTTTCAGCAACAGCTTAAACGCCATTTGCGCCATTTGTCGCAATGTCGTCGTCGCAATTT 3027

QY 563 CATTGAGCTTCGCGACCATATCAATATTTGTTTCAAGTTTCAACATGCCCTCAGACAGGACT 622  
 DB 3028 CATTCAATAAGTGTAGGCTGTGAGTTGCTTTCGTAAGCAATTTTCGGCGTCAGACCA 3087

QY 623 TACCTGGCTGGC-----CGGGAAACCTCTCGGAATTCATCGAGATTTTGT 670  
 DB 3088 GAACCTTTATTTCCCTTTCACGATGTCATCGGAAATCTTGGGAAATCCGCTTGAGATTTTC 3147

QY 671 CCGTGACGAGCGCTGCGCAAGTGTGAGAAAGCAATGACGCAAGCAATTTGTCGAC 730  
 DB 3148 CGGTGAGCAATTCCTGAGCAGAGGATTAAGGCAATACAGCCCGCTTATTTTGA 3207

QY 731 CTGACTGCAACAGTTCTCACCGTCAATGCACTGGTCAAGGTTATCGGCGGCTACTCTCTTGTCT 790  
 DB 3313 TTTCGCGCTGCTGCGGAGTA 3334

QY 791 AATAGCTTGGCTGATGAATCAGAAAGCGGCGAGCCCTCTCCGCGCATGAATCAGCCGCT 850  
Db 3253 TGTACGAAGGTTGATGAATTAACAGCGGAATTTTCCACTCGCGCAGCAACTCAACCAATTI 3312  
QY 851 CCGCTGTGAGCTCTGGACCGTA 872  
Db 3313 TTTGGCTCCGCTCTGGCGAGTA 3334

## RESULT 3

US-08-921-177-20  
; Sequence 20, Application US/08921177  
; Patent No. 5798211  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/921,177  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5541 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-625  
US-08-921-177-20

Query Match 8.2%; Score 71.6; DB 1; Length 5541;  
Best Local Similarity 49.4%; Pred. No. 1.7e-11;  
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;  
QY 203 CATTGTGACGCGCTGTCAGCCCAAGGTGGCGCTGGATCACTGCTGGTGGAG 262  
Db 2687 CATTATCCATCTGACCTTTGTGCTCTCTGCATTTGTGATGGGCTATCTTGACCGACTGAC 2746  
QY 263 TTGTGATGACCGCATTTGGCCATCAAACTGATTTGATGGTTAGTTTCGCGGTTTGG 322  
Db 2747 TCGTCATATCACTGATCTTATGCGGGCGGGTTCTCGCGCCGGTTATTAACAGGTCAAT 2806  
QY 323 AATCGGTGGCCTTCGCCCCAAATGTTGATGCCGCGCTGCTGGGAAATCTCATCGATCGCCT 382

Db 2807 TATCGGAAGACGCTGCCACAGATTGCTCGCATCGCGATATGCTGATCAATCTGCG 2866  
QY 383 CGAATCTCGGCTCAGAAAACCTCAAGTTTGTGAGTGAATCAAGGCTGTTGTCCAGTGTCT 442  
Db 2867 CCAGCTCTCCGCTGCTGTAATGTCAGATTATTCAGCGCTGTCAGCTTCTCCTCAAGTTGTC 2926  
QY 443 CAACTGACGAAGCACCAATCAATGCACTGCTCAGCGTATCCGCGCGTACTCTCCTTGTCT 502  
Db 2927 CGCGCGC-TGGCACCACATCAATACGAGCTCAGCGATCA-----T 2967  
QY 503 CGCGCAGCACCATGCAAGCGCATCTCGCAAGTGAATCTCCCGCGTTCCTTGGGCGATGT 562  
Db 2968 CTTTCAGCAACCAAGCTTAACGCCATTTCGCGCATTTGTTGTCAGCTGCTGTGCCATTT 3027  
QY 563 CATTGAGCTTCGGGACCATATCAATATTGTTCAAGTTCAGCTTCAACATGCCCTCAGACAGGACT 622  
Db 3028 CATTCAATAAGTGTAGGCTGTTGAGGTTTCGGTAAAGCATTTTCGGCGCTCAGACCCAC 3087  
QY 623 TACCTTGGCTGGC-----GCGGGAACCTCTCTGGAATTCATCGAGATATTTGT 670  
Db 3088 GAATTTATTCCTTCAGCATGCATCCGTAATCTTGGGAATGCCGTTGAGATATTTTC 3147  
QY 671 CGGTGAGCAGGCTTCGCAAGTGGTGAGAAAGCAATGACGCCAAGACCATTTGTTGGCAG 730  
Db 3148 CGGTGAGCAATCCCTGAGCCAGAGGAGTAAGAGCAATACAGCCCGCGTTATTTGCA 3207  
QY 731 CTGACTGCAACAAGTTCTCACCGTCTATCCCGCGTTCCTCCACCCCAACGATTAATGATGG 790  
Db 3208 GGGTATCCAGCA-----GCGCGCTTTTATCCACCCAGCGGTTTCAGTAAAT 3252  
QY 791 AATAGCTTGGCTGATGAATCAGAACGCGGCGCAGCCCTCTCCGCGCATGAATCAGCGGCT 850  
Db 3253 TGTACGAAGTTGATGAATTAACAGCGGAATTTTCACTTCGCGCAGCAACTCAACCATTT 3312  
QY 851 CCGCTGTGAGCTCTCGACCGTA 872  
Db 3313 TTTGGCTCCGCTCTGGCGAGTA 3334

## RESULT 4

US-08-362-577C-20  
; Sequence 20, Application US/08362577C  
; Patent No. 5807673  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Tsuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/362,577C  
; FILING DATE: 27-MAR-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300

TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5541 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORGANISM: Escherichia coli  
STRAIN: Clinical Isolate EC-625  
US-08-362-577C-20

Query Match 8.2%; Score 71.6; DB 1; Length 5541;  
Best Local Similarity 49.4%; Pred. No. 1.7e-11;  
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;

QY 203 CATTGTGACGGCCGCTGTCAGCCGCCAAGGTGTGGCGCTGGATCAACGTGCTGGCGAG 262  
DB CATTATCCATCTGACGTTTGTGCTCTCTGCAATTTGTGATGGGCTATCTTGACCGACTGAC 2746  
QY 263 TTGTGATGACCGCATGGCCATCAAACTGATGTTGATGGTTAGTTTTCGCGGGTTTGG 322  
DB TCGTCAATCACTGATCTTATGCGGCGGTTCTCGCGCCGTTATTAAACAGGTCATT 2806  
QY 323 AATCGGTGGCTTTCGCCAAATGTTGATGCGCGCTGTTGGGAAATCTCATGATCGCCT 382  
DB TATCGGAAGACGCTGTCACAGATTTCAGCTCGCCATCGCGGATATGCTGATCAATCTGCG 2866  
QY 383 CCAACTCGGCGTCAGAAACTCCAAGTTGTTGATGTAATCAAGCTGTTCTCCAGCTGCT 442  
DB CAGCTCTCTCGGTGCTAAATGTCAGATTATTCAGCGCTTCACGTTCTCTCAAGTTGTC 2926  
QY 443 CAACTGACGAAGCACCATCAATGCACTGCTCAGGTCATCCGCGCGTACTCTCTTGTCT 502  
DB CGCGCGC-TGGCACCATCAATACCGAGTCACGGATCA-----T 2967  
QY 503 CGCGCAGCACCATGCAAGCGCCATCTGCGCAAGTGAAGTTCCTCGCGGATGT 562  
DB CTTTCAGCAACCAAGCTTAACGCCATTGCGCCATTGTTGTCACGCTGCTGTGCCATT 3027  
QY 563 CATTGAGCTTGGGACCATATCAATATTTGATGTTCAAGTTCACATGCCCTCAGACGGACT 622  
DB CATTCAATAAGTGTAGGCTGTGAGGTTGCTTCGGTAAGCATTTTCGGGTCAGACCC 3087  
QY 623 TACCTGTGCTGC-----GCGGGAACCTCTGGAATTCATCGAGATTTGT 670  
DB GAATTTATTCCTTCAGATGCTCCGTGAATCTTGGGAAATGCGTTGAGATTTTC 3147  
QY 671 CCGTGAGCAGCGCTTGGCAAGTGTGAGAAAGCAATGACGCCAAGACCATTTGTTGGAG 730  
DB CGGTGAGCAATCTCTGAGCAGAGAGTAAAGGCAATACAGCCACGCGCTTATTTTGA 3207  
QY 731 CTGACTGCAACAGTTCTCACCGTCATCGCGGCTTCCTCCACCCAGATTAATGAG 790  
DB GGGTATCCAGCA-----GCGCGCTTTTATTCACCCAGCGGTTCAGTAAAT 3252  
QY 791 AATAGCTTGGCTGATGAATCAGAGCGGCGAGCCCTCTCCGCCATGAATCAGCGGCT 850  
DB TGTACGAGGTTGATGAATTAACGCGGAATTTTCCATCTCGCGGAGCACTCAACCATTT 3312  
QY 851 CCGCTGTGAGCTTGGACCGTA 872  
DB TTTGCTCGGCTCTGGCGAGTA 3334

RESULT 5  
US-08-920-828-20  
Sequence 20, Application US/08920828  
Patent No. 5853998  
GENERAL INFORMATION:  
APPLICANT: Ohno, Tsuneya

APPLICANT: Matsuhisa, Akio  
APPLICANT: Uehara, Hirotsugu  
APPLICANT: Eda, Soji  
TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/920,828  
FILING DATE: 29-AUG-1997  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,577  
FILING DATE: 27-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rin-Laures, Li-Hsien  
REGISTRATION NUMBER: 33,547  
REFERENCE/DOCKET NUMBER: 19036/32420  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/474-6300  
TELEFAX: 312/474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5541 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: Genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Escherichia coli  
STRAIN: Clinical Isolate EC-625  
US-08-920-828-20

Query Match 8.2%; Score 71.6; DB 2; Length 5541;  
Best Local Similarity 49.4%; Pred. No. 1.7e-11;  
Matches 337; Conservative 0; Mismatches 299; Indels 46; Gaps 4;

QY 203 CATTGTGACGGCCGCTGTCAGCCGCCAAGGTGTGGCGCTGGATCAACGTGCTGGCGAG 262  
DB CATTATCCATCTGACGTTTGTGCTCTCTGCAATTTGTGATGGGCTATCTTGACCGACTGAC 2746  
QY 263 TTGTGATGACCGCATGGCCATCAAACTGATGTTGATGGTTAGTTTTCGCGGGTTTGG 322  
DB TCGTCAATCACTGATCTTATGCGGCGGTTCTCGCGCCGTTATTAAACAGGTCATT 2806  
QY 323 AATCGGTGGCTTTCGCCAAATGTTGATGCGCGCTGTTGGGAAATCTCATGATCGCCT 382  
DB TATCGGAAGACGCTGTCACAGATTTCAGCTCGCCATCGCGGATATGCTGATCAATCTGCG 2866  
QY 383 CCAACTCGGCGTCAGAAACTCCAAGTTGTTGATGTAATCAAGCTGTTCTCCAGCTGCT 442  
DB CAGCTCTCTCGGTGCTAAATGTCAGATTATTCAGCGCTTCACGTTCTCTCAAGTTGTC 2926  
QY 443 CAACTGACGAAGCACCATCAATGCACTGCTCAGGTCATCCGCGCGTACTCTCTTGTCT 502  
DB CGCGCGC-TGGCACCATCAATACCGAGTCACGGATCA-----T 2967  
QY 503 CGCGCAGCACCATGCAAGCGCCATCTGCGCAAGTGAAGTTCCTCGCGGATGT 562  
DB CTTTCAGCAACCAAGCTTAACGCCATTGCGCCATTGTTGTCACGCTGCTGTGCCATT 3027  
QY 563 CATTGAGCTTGGGACCATATCAATATTTGTTCAAGTTTCAACATGCTCCCTCAGACGGACT 622

Db	3028	CATTCAATAAGTGTAGGCTGTTGAGGTTGGCTTCGGTAAGACATTTTCGGCGTCAGACCAC	3087
Qy	623	TACCCCTGGCTGGC-----GCGGGAACCCCTCTGGGAATTCATCCGAGATATTGT	670
Db	3088	GAACTTTATTCCTTCACGATCATCCGTGATCTTCGGGAATGCCGTTGAGATATTTTC	3147
Qy	671	CCGTGAGCAGGCCCTCGCGAAGTGTGTGAGAAAGCAATGACGCCAAGACCAATTGTTGGCAG	730
Db	3148	CGGTCAGCAATCCCTGAGCCAGAGGAGTAAGAGCAATACAGCCACGCGGTTATTTTGCA	3207
Qy	731	CTGACTGCAACAAAGTTCTCACCGTCATCGCCGGTTCTCTCACCCACAGGATTAAATGATGG	790
Db	3208	GGGTATCCAGCA-----GGCCGCTTTTATCCACCACGCGGTTCAGTAAAT	3252
Qy	791	AATAGCTTGGCTGATGAATCAGAAGCGGCGACGCCCTCCTCGGCCATGAATCTCAGCCGCCT	850
Db	3253	TGTACGAAGGTTGATGAATTAACAGCGGNAATTTTCCATCTCGGCGAGCAATCTCAACCATTT	3312
Qy	851	CCGCTGTGAGCTCTGACCGTA	872
Db	3313	TTTTCGCTCGCCTCTGGCGAGTA	3334

## RESULT 6

```

US-08-390-878-17/c
; Sequence 17, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; STREET: Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15239 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-390-878-17

```

Qy	86	TTGTGTTTATACGGCGGCGTCCGCGCGCAATACGCGCAGACCGGACGCTGATTTTCGCCG	145
Db	4732	TCGTGTTTGTCTAGGCGCGCTGGCCAACGAGCACACGCA---CCACGCGTGGCTGTTCCGCC	4676
Qy	146	CTCGCGGCTTCGCGGCAAGCCTGATCTTGTTCCCGCTGGTGGGTTTCGGCGCAGCAGCAT	205
Db	4675	TCGCGGCGGTACAGCAGATGCGGTATGGTTGCGCACCCCTCGGGTTCGGAGCCGCCCGGT	4616
Qy	206	TGTACGCGCGCTCTTCAGGCCCAAGCTGTGGCGCTGGATCAACGTCGTCTGTGCAGTTG	265
Db	4615	TGCGCGGCTGTTTACCAACACCCCGGCTCTGGAGAACTCTTCAGCGGCTGATCGCGTCA	4556
Qy	266	TGATGACCGCATTTGGCCATCAAACTGATGTTGA	298
Db	4555	TGATGGTTGGCTGGGAATCTCGCTGACCGTGA	4523

## RESULT. T 7

```

US-09-103-840A-2
US-09-103-840A-2
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R. A. M.
APPLICANT: FRASER, John C.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
TITLE OF INVENTION: TUBERCULOSIS
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

```

## RESULT 8

RESULI 8  
US-09-103-840A-1  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328

; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 6.5%; Score 57; DB 4; Length 4411529;  
Best Local Similarity 53.1%; Pred. No. 9.1e-06;  
Matches 145; Conservative 0; Mismatches 125; Indels 3; Gaps 1;

QY 26 AGCCCATGTTGATGGCAATCGTGTGACCTGGTGTGAACCGAATGCGTATTTCGACCGT 85  
Db 2230334 AGTCTCTGTGACCTGTGGGCAATCAGTTCTCAACCCACACGTCCTACCTCGACACCG 2230393

QY 86 TTGTGTTTATCGGCGCGGTGCGCGCGCAATACGCGGACACACCGGACGGTGGATTTCGCCG 145  
Db 2230394 TCGTGTGTAGCGCGCTGGCCACAGCACAGCA---CCAGCGCTGGCTGTTGCGCC 2230450

QY 146 CTGCGCGGTTCGCGGCAAGCCTGATCTGTTCCCGTGTGGTGGTTCGCGGACAGCAT 205  
Db 2230451 TCGGCGCGGTTCACAGCAGTGGGTGTTGCGCACCTCTCGGGTTCGAGCGCGCGGT 2230510

QY 206 TGTCACCGCGCTGTCAGCCCAAGGTGGCGCTGGATCAAGTCTGTCGCACTTG 265  
Db 2230511 TCGCGCGGTGTTTCACCAACCGCGCTGGTGGAGATCTTCGACGCGCTGATCGCGGTCA 2230570

QY 266 TGATGACCGCATTTGGCCATCAAACTGATGTTGA 298  
Db 2230571 TGATGTTGCGCTGGCAATCTGCTGACCGTGA 2230603

RESULT 9  
US-08-492-027A-9/c  
; Sequence 9, Application US/08492027A  
; Patent No. 5912333  
; GENERAL INFORMATION:  
; APPLICANT: Suzuki, Shoichi  
; APPLICANT: Burnell, James N  
; TITLE OF INVENTION: DNA Encoding Carbonic Anhydrase  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch and Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/492,027A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Murphy Jr, Gerald M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 0760-206

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1935 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: other nucleic acid  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1638  
US-08-492-027A-9

Query Match 4.5%; Score 39.6; DB 2; Length 1935;  
Best Local Similarity 49.1%; Pred. No. 0.038;  
Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 222 CAGCCCAAGGTGTGGCGCTGGATCAACGTGCTGGGAGTTGTGATGACCGCATTTGGC 281  
Db 1476 CGGACCGAGGGGTGCTCTTCTTTACCTTCATCTTGGCAATGAAGCCGATCTTAACCCA 1417

QY 282 CATCAAACTGATGTTGATCGGTTAGTTTTCGGGGTTTGGATCGTGGCCTTCGCCCA 341  
Db 1416 GTCTCGACGAAGTGAAGGTGAGCTGCGCGGTTCTTTGAGGGAGAGAGAGCGCCCT 1357

QY 342 AATGTTGATGCGCGGCTCGTGGGAATCTCATGATCGCTCCAACTCGCGCGTCAGAAA 401  
Db 1356 GATGCCACCGCAGCAGTATGGCCATGACGAGGACTCCACCTTGAGGGCGCACAC 1297

QY 402 CTCCAAAGTTGTTGAGTGAATCAAGGCTGTGTGCC 435  
Db 1296 AGCGTACTCGATGGCGGACCGGATGCGCGGTAC 1263

RESULT 10  
US-09-103-840A-2/c  
; Sequence 2, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TITLE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 4403765  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; FEATURE:  
; OTHER INFORMATION: CDC 1551  
; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
; OTHER INFORMATION: represent a, t, c or g  
US-09-103-840A-2

Query Match 4.3%; Score 37.6; DB 4; Length 4403765;  
Best Local Similarity 47.5%; Pred. No. 5.4;  
Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 25 AAGCCCATGTTGATGCAATCGTGTGCTGACCTGTTGAACCCGAATGCGTATTTGGAGCG 84  
Db 720443 AAGGACCTGCGCTGGCGCTGCGAGCGGCGGCGAGCTGCCCGGCCCGCGGTGGT 720384

QY 85 TTTGTGTTTATCGGCGCGTTCGCGCGCAATACGCGGACACCGGACGTTGATTTTCGCC 144  
Db 720383 GGCTGCTGTTTACCTTGCAACGGCGGCGGAGATGTTTCGGGGTCCACGACGAC 720324

QY 145 GCTGGCGGCTTCCGGCAAGCTGATCTGTTCCGCTGGTGGTTTCGGCGCAGCA 204  
DB 720323 CGCTCGACGATCGAGGACCTGCTGGCGGGATTCGCTGGCGGTTCTTCGCGCGGG 720264  
QY 205 TTGTACGCCCGCTGTCAGCCCAAGTGTGGCGCTGATCAAGCTGCTGTGGC 260  
DB 720263 GAGATCGGCCCGGTCGGCGGCCACACGCTGTGCACGGGTTTACCGGCTCATGGC 720208

RESULT 11  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; FILE OF INVENTION: TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H37Rv  
US-09-103-840A-1

Query Match 4.3%; Score 37.6; DB 4; Length 4411529;  
Best Local Similarity 47.5%; Pred. No. 5.4;  
Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 25 AAGCCCATGTTGATGCAATCGTCTGACCTGGTGTGAACCCGAATGCGTATTGGACGG 84  
DB 719010 AAGGACCTGGCGCTGGCGTGGAGGGGGCGCGAGCTGCCGGGCCCCGGTGGT 718951  
QY 85 TTGTGTATTATCGCGGGGCTGGCGCGGATACGGCGACACCGGACGGTGGATTTTCGCC 144  
DB 718950 GGGCTGCTGTTCACCTGCAACGGCGCGGACGAGAAATGTTGGGGGTCAACGACCAAGC 718891  
QY 145 GCTGGCGGCTTCCGGCAAGCTGATCTGTTCCGCTGGTGGTTTCGGCGCAGCA 204  
DB 718890 CGCTCGACGATCGAGGACCTGCTGGCGGGATTCGCTGGCGGGTTTTCGCGCGGG 718831  
QY 205 TTGTACGCCCGCTGTCAGCCCAAGTGTGGCGCTGGATCAACGCTGCTGTGGC 260  
DB 718830 GAGATCGGCCCGGTCGGCGGCCACACGCTGTGCACGGGTTTACCGGCTCATGGC 718775

RESULT 12  
US-08-776-246-1/c  
; Sequence 1, Application US/08776246  
; Patent No. 6004550  
; GENERAL INFORMATION:  
; APPLICANT: Springer, Caroline J  
; APPLICANT: Marais, Richard  
; TITLE OF INVENTION: Intracellular expression of carboxypeptidase  
; TITLE OF INVENTION: G2 in enzyme produg therapy  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderyhe  
; STREET: 1100 No. 6004550th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,246  
FILING DATE: 30-JAN-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/01783  
FILING DATE: 27-JUL-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9415167.7  
FILING DATE: 27-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Arthur R. Crawford  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 620-19  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2048 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
US-08-776-246-1

Query Match 3.9%; Score 34.4; DB 3; Length 2048;  
Best Local Similarity 48.0%; Pred. No. 1.4;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;  
QY 461 TCAATGCACTGGTCAAGTATCGCGCCGCTACTCTCTTGTCTGGCGAGCAGCCATGCAA 520  
DB 1950 TCGATGCACTGGTGGCGGCTGGGCTCGTCTAGGAGAGGCGCGGATCCGGCCCT 1891  
QY 521 GGGCCATTCGCCAAGTACTGCCCGCTTCTCTGGCGAGTGTCTTGGCTTGGCGAGCA 580  
DB 1890 GGGCCAGCGGCTGAGCCACTGCTTCTCTGCTGGCGGTTGGCTAGCGCATGAGGATCG 1831  
QY 581 TATCATATTTTTCAGTTCAATGATGCCCTCAGACAGGAGACTTACCTGGCTGGCGGG 640  
DB 1830 CGTTGACGGCGGCGAGTTGGTTCAGCTGATGGCGGTGTGTGGCGGCTGCTGGCGGCGCA 1771  
QY 641 AACCTCTCGAATTCATCGAGAT 664  
DB 1770 TCTCTTCCAGCAGCAGCGCAGGT 1747

RESULT 13  
US-08-776-251-1/c  
; Sequence 1, Application US/08776251  
; Patent No. 6025340  
; GENERAL INFORMATION:  
; APPLICANT: Springer, Caroline J  
; APPLICANT: Marais, Richard  
; TITLE OF INVENTION: Surface expression of enzyme in gene directed produg therapy  
; NUMBER OF SEQUENCES: 27  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon & Vanderyhe  
; STREET: 1100 No. 6025340th Glebe Road, 8th Floor  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: USA  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
; CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,251  
FILING DATE: 31-JAN-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/01782  
FILING DATE: 27-JUL-1995

PRIOR APPLICATION DATA: GB 9415167.7  
FILING DATE: 27-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Arthur R. Crawford  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 620-20  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2048 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
US-08-776-251-1

Query Match 3.9%; Score 34.4; DB 3; Length 2048;  
Best Local Similarity 48.0%; Pred. No. 1.4;  
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 461 TCAATGCACTGGTCACGGTATCCGCGCTACTCTCTCTGTCGCGCAGCACCCTGCA 520  
DB 1950 TCGATCATCGTCGCGCTCGGCTCGGTTCAGGCAGAGGCGCGGATCGGGCCT 1891  
QY 521 GCGCATCTGCGCAAGTACTGCCCGGTTCTCTGGCGATGTCTATGAGCTTGGGACCA 580  
DB 1890 GCGCAGCGGCTCAGCACTGCTTCTCTGCTCGCGGTTGCCGTAGCGATGAGGATGG 1831  
QY 581 TATCAATATTGTTTACGTTTCAACATGCCCTCAGACAGGACTTACCTGCTGGCGGGG 640  
DB 1830 CGTTGACGGGCGAGTTGGTTCACGCTGATGGCGGTGCTGTGCGCGCTCGCGGCGCGA 1771  
QY 641 AACCTCTGGAATTCATCGAGAT 664  
DB 1770 TCTCTCAGCACCAGCGCGAGGT 1747

RESULT 14  
US-09-434-288-7  
Sequence 7, Application US/09434288  
Patent No. 6303767  
GENERAL INFORMATION:  
APPLICANT: Betlach C., Melanie  
APPLICANT: McDaniel, Robert  
TITLE OF INVENTION: POLYPEPTIDE SYNTHASE ENZYMES AND RECOMBINANT DNA  
TITLE OF INVENTION: CONSTRUCTS THEREFOR  
FILE REFERENCE: 30062-20030.00  
CURRENT APPLICATION NUMBER: US/09/434,288  
CURRENT FILING DATE: 1999-11-05  
PRIOR APPLICATION NUMBER: 60/107,093  
PRIOR FILING DATE: 1998-11-05  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 1681  
TYPE: DNA  
ORGANISM: Streptomyces narbonensis  
US-09-434-288-7

Query Match 3.8%; Score 33.6; DB 4; Length 1681;  
Best Local Similarity 49.4%; Pred. No. 2.2; Indels 0; Gaps 0;  
Matches 87; Conservative 0; Mismatches 89;

QY 115 TACGGGACACCGGACGGTGGATTTTCGCGCTGGCGGTTTCGCGGCAAGCTGATCTGG 174  
DB 1493 TACTGCGGTCGCGGGGTTCTTCCAGCAGCGGGGAGTTTCGACGCGGACTTCTTCGGG 1552  
QY 175 TTCCGCTGGTGGGTTTCGGGCGGAGCAGATTTCACGCGCGCTGTTCAGCCCAAGTG 234  
DB 1553 ATCTGCGCGCGAGGCGCTCGCATGACCCGCGAGCAGCGGCTGTCTCCACACCGCG 1612  
QY 235 TGGCGCTGGATCAACGTCGTGTGGCAGTTGTGATGACCGGATTTGGCCATCAAACT 290

DB 1613 TGGGAGGCGATCGAGCAGCGGGGATCGACCGAGCTGAAGGCGAGGGCCT 1668

RESULT 15  
US-07-945-283-1  
Sequence 1, Application US/07945283  
Patent No. 5352596  
GENERAL INFORMATION:  
APPLICANT: Cheung, Andrew K.  
APPLICANT: Wesley, Ronald D.  
TITLE OF INVENTION: Pseudorabies Virus Deletion Mutants  
TITLE OF INVENTION: Involving The BP0 and LLT Genes  
NUMBER OF SEQUENCES: 7.  
CORRESPONDENCE ADDRESS:  
ADDRESS: Curtis P. Ribando  
STREET: 1815 No. 5352596th University Street  
CITY: Peoria  
STATE: IL  
COUNTRY: USA  
ZIP: 61604  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/945,283  
FILING DATE: 19920911  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Ribando, Curtis P.  
REGISTRATION NUMBER: 27976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 309-685-4011 ext. 513  
TELEFAX: 309-685-4128  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8438 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Pseudorabies virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 622..6495  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1099, "g")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1267, "t")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1381, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(1566, "c")  
FEATURE:  
NAME/KEY: variation  
LOCATION: replace(7010, "g")  
US-07-945-283-1

Query Match 3.8%; Score 33.6; DB 1; Length 8438;  
Best Local Similarity 51.3%; Pred. No. 4.7; Indels 0; Gaps 0;  
Matches 78; Conservative 0; Mismatches 74;

QY 337 GCCCAATGTTGATCGCGGCTCGTGGGAAATCTCATCGATCGCTCCCAACTCGGGGTCA 396  
DB 6434 GCCCGACGGGGACCGCGGCGCGGGGCTCGTCTCTCTCTCTCTCTCTCTCTCTCG 6493

Qy 397 GAAACTCCAAGTTGTTGAGTGAATCAAGGCTGTTGTCCAGCTGCTCAACTGACGAAGCA 456  
Db 6494 TCTAGCACACGATCTCGCCCGAGCCCGCGGGGGTGCCTGCTGCTGGGGCCGAGGA 6553  
Qy 457 CCAATCAATGCACTGCTGTCACGGTATCCGGCC 488  
Db 6554 GGACGGGGCGGCTCGTGGCTCCGGCCCGCGC 6585

Search completed: April 27, 2003, 07:20:35  
Job time : 12814.2 secs





GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:21:23 ; Search time 19 Seconds  
(without alignments)  
1194.091 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVIMEIFITGLLGASLLLS.....INVVVAVVMTALAIKLMLMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 73:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	350	29.4	211	1 QOEC5A	hypothetical 23K p
2	347	29.1	211	2 AH0874	probable membrane
3	343	28.8	211	2 B91103	hypothetical prote
4	340	28.5	211	2 E85948	hypothetical prote
5	335	28.1	205	2 AG0112	probable LyseE type
6	335	27.9	200	2 D83100	probable membrane
7	326.5	27.4	202	2 C97472	probable membrane
8	326.5	27.4	202	2 AG2690	Lyse family transp
9	319	26.8	201	2 C70744	hypothetical prote
10	318	26.7	199	2 H70756	hypothetical prote
11	302.5	25.4	204	2 AD3411	transporter, lyse
12	245.5	20.6	211	2 B82318	Lyse/Ygga family p
13	238	20.0	211	2 S57940	Ygga protein homol
14	176.5	14.8	220	2 D69838	conserved hypotet
15	173.5	14.6	210	2 F64609	conserved hypotet
16	170.5	14.3	210	2 B71906	hypothetical prote
17	164	13.8	205	2 B89854	conserved hypotet
18	154.5	13.0	205	2 C97219	uncharacterized co
19	153.5	12.9	213	2 E97789	hypothetical prote
20	151.5	12.7	208	1 B69066	hypothetical prote
21	148	12.4	200	2 G83703	conserved hypotet
22	136.5	11.5	213	2 F83444	hypothetical prote
23	132	11.1	210	2 E87252	hypothetical prote
24	130	10.9	210	2 D84016	efflux protein, Ly
25	120.5	10.1	208	2 G87305	hypothetical prote
26	119	10.0	620	2 F84449	efflux protein, Ly
27	117.5	9.9	197	2 B83280	potassium uptake p
28	115.5	9.7	216	2 AH3203	hypothetical prote
29	114.5	9.6	208	2 G84086	RhtB family transp
					dihydrodipicolinat

30	113	9.5	222	2 B87264	efflux protein, Ly
31	113	9.5	235	2 F97679	hypothetical prote
32	113	9.5	235	2 AD2904	RhtB family transp
33	111	9.3	212	2 AD0714	probable membrane
34	109.5	9.2	218	2 A96009	probable amino aci
35	109.5	9.2	450	2 E72615	probable hlyE AP81
36	108.5	9.1	203	2 AB2683	RhtB family transp
37	108.5	9.1	239	2 H97464	conserved hypotet
38	107.5	9.0	284	2 H69277	branched-chain ami
39	107	9.0	210	2 G83082	hypothetical prote
40	105	8.8	206	2 A98223	threonine efflux p
41	104.5	8.8	349	2 G98273	iron(III) dicitrat
42	104.5	8.8	349	2 AG3010	hypothetical prote
43	104	8.7	205	2 B83356	conserved hypotet
44	102	8.6	206	2 AH0466	threonine efflux p
45	102	8.6	212	2 T43921	yfud protein limpo

## ALIGNMENTS

### RESULT 1

#### QOEC5A

hypothetical 23K protein (sbm-fba intergenic region) - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 31-Dec-1990 #sequence\_revision 17-Oct-1997 #text\_change 01-Mar-2002  
C:Accession: B65077; S04736

R:Blatner, P.R.; Flunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: B65077

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-211 <BLAT>

A:Cross-references: GB:AE000375; GB:U00096; NID:q1789282; PIDN:AAC75960.1; PID:q1789290,

A:Experimental source: strain K-12, substrain MG1655

R:Alfouf, P.R.; Pernam, R.N.

Mol. Microbiol. 3, 723-732, 1989

A:Title: Identification, molecular cloning and sequence analysis of a gene cluster encod

hyde 3-phosphate dehydrogenase of Escherichia coli.

A:Reference number: S04730; MUID:89313302; PMID:2546007

A:Accession: S04736

A:Molecule type: DNA

A:Residues: 15-211 <ALE>

A:Cross-references: EMBL:X14436; NID:g41417; PIDN:CAA32607.1; PID:g41425

C:Genetics:

A:Gene: yggA

A:Map position: 63 min

C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)

Query Match 29.4%; Score 350; DB 1; Length 211;  
Best Local Similarity 35.7%; Pred. No. 1.6e-24;  
Matches 81; Conservative 37; Mismatches 75; Indels 34; Gaps 4;

QY 7 FITGILLGASLLISGPQNVIVKQIKREGIAVLLVCLISDVFLFIAGTIGVDLLSNA 66

Db 5 YFGGLGALGAAMTILPLGPQNAFVNMQIRQYHIMIALCAISDLVLCAGIFGGSALLMQ 64

QY 67 APVLIDIMEWGGIAYLLWFAVMAAKDMTKVEAPQIIEETPTVPDDTFLGGSAVATDT 126

Db 65 SPWLLALVTWGGVAVLLMTWFGAFKATMSSNIEL----- 98

QY 127 RNRVRVSVDRQVRVWVPMALVLTWLNPNAYLDAFVIGGVGAQYG-DTRGWIFAAG 185

Db 99 -----ASAEVWKQGRW-KLIATMLAVTWNPHVLDTFVVLGSLGQLDVEPKRW-FALG 151

QY 186 AFAASLIWPLVGFCAALSRPLSPKVRWVNVVAVVMTALAIAKL 232

Db 152 TTSASFLNFFGLLAAWLAAPRLRTAKAQRILNVVGVCMWFIALQL 198

[illegible]

A:Gene: YPO0918  
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)  
Query Match 28.1%; Score 335; DB 2; Length 205;  
Best Local Similarity 33.1%; Pred. No. 3.6e-23;  
Matches 80; Conservative 44; Mismatches 70; Indels 48; Gaps 6;  
QY 3 IMEIFTGLLGASLLSIPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVDL 62  
DB 1 MLAVYLHGFTLSAAILPLGPQNVFVWVQKIKQHLMSALCALSDIILICAGIFGSA 60  
QY 63 LSNAAPVLDIMRWGGIAYLLWF---AVMAAKDAMTKVAPQIIEETEPTVDDTPLGG 119  
DB 61 LLSRSPLLLALVTWGGVAFPLMYGWGALMAA-----WRG 94  
QY 120 SAVATD-----TRNRVRVESVDQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGVGCAQ-Y 174  
DB 95 DGVASATSVTQGRWILLVT-----LAVTLNPNHYLDTFVVLGSLGQLL 141  
QY 175 GDTGRWIFPAAGAPASLIWFLPVGGAALSRPLSSPKVWRWVNVVAVMTALAIKML 234  
DB 142 PDIRPW-FALGAVTASIVWFFALALLAALWLSPLNRPVQRIINLFPVGMGFIAQLAR 200  
QY 235 MG 236  
DB 201 QG 202  
RESULT 6  
D83100  
Probable transporter PA4365 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: D83100  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: AB2950; MUID:20437337; PMID:10984043  
A:Accession: D83100  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-200 <STO>  
A:Cross-references: GB:AE004852; GB:AE004091; NID:g9950587; PIDN:AAG07753.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4365  
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)  
Query Match 27.9%; Score 332; DB 2; Length 200;  
Best Local Similarity 33.5%; Pred. No. 6.5e-23;  
Matches 77; Conservative 42; Mismatches 77; Indels 34; Gaps 4;  
QY 7 FITGLLGASLLSIPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVLLSNA 66  
DB 5 YLNGILVAAGLIAIGAQNAPFVLAQSLREHLSVAALCVFCDAVLVSLGVFLAKLLE 64  
QY 67 APVLDIMRWGGIAYLLWFPAVMAAKDAMTKVAPQIIEETEPTVDDTPLGGSAVATDT 126  
DB 65 NPTLIAIARWGGIATVTVGLKALLRALR-----PDAL---GNAETGP 105  
QY 127 RNRVRVESVDQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGVGCAQYGDGTGRWIFPAAG 186  
DB 106 RSRKAV-----LLAALATVTLNPNHYLDTFVVLGSLGQAQAPG--AYALGA 150  
QY 187 PAASLIWFLPVGGAALSRPLSSPKVWRWVNVVAVMTALAIKMLMG 236  
DB 151 ASASLWFFPALAALGAALPWLARPATWELLDLIMVAAMLMGMAQLIFRG 200  
RESULT 7  
C97472

probable membrane transport protein. (AL357613) [imported] - Agrobacterium tumefaciens (C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: C97472  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: C97472  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-202 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK6732.1; PID:gl5155924; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C 1690  
A:Map position: circular chromosome  
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)  
Query Match 27.4%; Score 326.5; DB 2; Length 202;  
Best Local Similarity 31.6%; Pred. No. 2.1e-22;  
Matches 73; Conservative 42; Mismatches 83; Indels 33; Gaps 4;  
QY 4 MEIFTGLLGASLLSIPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVLL 63  
DB 3 IQIIFTGLTMGLSLIVAIGAQNAPFVLAQSLARSHVFAVCATCAISDALLIMVGVFGFQRI 62  
QY 64 SNAAPVLDIMRWGGIAYLLWFPAVMAAKDAMTKVAPQIIEETEPTVDDTPLGSAVA 123  
DB 63 SAIMPALDPIIMRYAGAAFLIYGAKSLSALRSS-EVLSVAERREAS----- 108  
QY 124 TDTNRVRVESVDQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGVGCAQYGDGTGRWIFA 183  
DB 109 -----LW-QTALCLALTPLNPHVLDITVLLGTISTQPPGFBEK-TFA 149  
QY 184 AGAPAAALIWFLPVGGAALSRPLSSPKVWRWVNVVAVMTALAIKML 234  
DB 150 AGAATGSLLPFSLGVGARMRLPIEKEFSARWILSGVIAITWTAIAFLVM 200  
RESULT 8  
AG2690  
LyseE family transporter lyseE [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002  
C:Accession: AG2690  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AG2690  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-202 <KUR>  
A:Cross-references: GB:AE008688; PIDN:AAL41941.1; PID:gl7739308; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: lyseE  
A:Map position: circular chromosome  
C:Superfamily: Escherichia coli hypothetical 23K protein (sbm-fba intergenic region)  
Query Match 27.4%; Score 326.5; DB 2; Length 202;  
Best Local Similarity 31.6%; Pred. No. 2.1e-22;  
Matches 73; Conservative 42; Mismatches 83; Indels 33; Gaps 4;  
QY 4 MEIFTGLLGASLLSIPQNVLVKQIKREGIAVLLVCLISDVFLFIAGTGLGVLL 63  
DB 3 IQIIFTGLTMGLSLIVAIGAQNAPFVLAQSLARSHVFAVCATCAISDALLIMVGVFGFQRI 62



[illegible]

A/Accession: F04005  
A/Status: preliminary; nucleic acid sequence not shown:

[illegible]



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:20:37 ; Search time 11 Seconds  
(without alignments)  
889.856 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVMEIFITGLLGALLS.....INWVAVVMTALAIKIMLMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1178	98.9	233	1 LYSE CORGL	P94633 corynebacte
2	350	29.4	211	1 YGGA ECOLI	P11667 escherichia
3	319	26.8	201	1 Y488 MYCTU	Q11154 mycobacteri
4	318	26.7	199	1 YJ86 MYCTU	Q10871 mycobacteri
5	310	26.0	206	1 YGGA AERSA	P70775 aeromonas s
6	254	21.3	225	1 YGGA AERHY	P52047 aeromonas h
7	105	8.8	206	1 RHTC ECOLI	P27849 escherichia
8	101	8.5	212	1 YBAS ECOLI	P17334 escherichia
9	93.5	7.9	452	1 PTCC ECOLI	P17334 escherichia
10	92.5	7.8	465	1 Y093 RHIME	O87394 rhizobium m
11	92	7.7	223	1 YAHN ECOLI	P75693 escherichia
12	91	7.6	206	1 RHTC ECOLI	Q91877 salmonella
13	89	7.5	944	1 CHS2 NEUCR	P30589 neurospora
14	88.5	7.4	3137	1 CA36 CHICK	P15989 gallus gall
15	88	7.4	206	1 RHTC SALT1	Q823b3 salmonella
16	87.5	7.3	443	1 DCUA HELPY	O25425 helicobacte
17	87	7.3	295	1 YBIF ECOLI	P74867 salmonella
18	86.5	7.3	195	1 YPIK ECOLI	P38101 escherichia
19	86	7.2	295	1 YBIF ECOLI	P36545 escherichia
20	86	7.2	351	1 MRAY METTH	O26830 methanococc
21	85	7.1	302	1 Y091 METJA	Q57556 methanococc
22	84.5	7.1	443	1 DCUA HELPY	Q921c0 helicobacte
23	84	7.1	482	1 YFIG BACSU	P54723 bacillus su
24	84	7.1	585	1 PHUB SALT1	O87656 salmonella
25	83	7.0	216	1 YBIF PSEAB	P38102 pseudomonas
26	82.5	6.9	3491	1 ERYI SACER	Q03131 saccharopol
27	81.5	6.8	253	1 CYSZ SALT1	Q824w3 salmonella
28	81.5	6.8	742	1 DHET ACEAC	P18278 acetobacter
29	81	6.8	622	1 KUP ECOLI	P30016 escherichia
30	81	6.8	676	1 CCMF RHIME	P45404 rhizobium m
31	80.5	6.8	253	1 CYSZ ECOLI	P12610 escherichia
32	80.5	6.8	253	1 CYSZ SALT1	P12673 salmonella
33	80	6.7	206	1 RHTB ECOLI	P27847 escherichia

34	80	6.7	394	1 YGAY ECO57	Q8x4v6 escherichia
35	80	6.7	461	1 CSBC BACSU	P46333 bacillus su
36	79.5	6.7	215	1 ORN MYCLE	O07708 mycobacteri
37	79.5	6.7	431	1 NQOE TETHH	Q56229 thermus the
38	79.5	6.7	459	1 DINF ECOLI	P28303 escherichia
39	79.5	6.7	479	1 XYLP LACPE	P96792 lactobacill
40	79	6.6	203	1 CHPE PSEAE	O87005 pseudomonas
41	79	6.6	593	1 MDLB ECOLI	P75706 escherichia
42	78.5	6.6	409	1 YM65 MYCTU	Q50697 mycobacteri
43	78.5	6.6	436	1 MYTH DEIRA	Q9rtpp deinococcus
44	78.5	6.6	457	1 AROP ECOLI	P15993 escherichia
45	78.5	6.6	457	1 AROP SALT1	Q91410 salmonella

## ALIGNMENTS

### RESULT 1

LYSE CORGL STANDARD; PRT; 233 AA.  
ID AC P94633;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Lysine exporter protein.  
GN LYSE OR CGL1262.  
OS Corynebacterium glutamicum (Brevibacterium flavum)  
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=1718;  
RN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RC STRAIN=97126810; PubMed=89711704;  
RX MEDLINE=97126810; PubMed=89711704;  
RA Vrljic M.M., Sahm H., Eggeling L.;  
RT "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum.";  
RL Mol. Microbiol. 22:815-826(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
RA Nakagawa S.;  
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";  
Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: INVOLVED IN THE EFFLUX OF EXCESS OF L-LYSINE. THIS IS NECESSARY TO CONTROL THE INTRACELLULAR L-LYSINE LEVEL.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X96471; CAA65324.2; -  
CC EMBL; AF005277; BAB98655.1; ALT\_INIT.  
CC InterPro; IPR001123; Lyse.  
CC InterPro; IPR004777; Lyse\_exporter.  
CC Pfam; PF01810; Lyse; 1.  
CC TIGRPFAM; TIGR00948; 2a75; 1.  
KW Transport; Transmembrane; Inner membrane.  
FT TRANSMEM 3 23 POTENTIAL.  
FT TRANSMEM 35 55 POTENTIAL.  
FT TRANSMEM 66 86 POTENTIAL.  
FT TRANSMEM 144 164 POTENTIAL.  
FT TRANSMEM 177 197 POTENTIAL.  
FT TRANSMEM 213 233 POTENTIAL.  
SQ SEQUENCE 233 AA; 25082 MW; F5FD9B1ACAD11D13 CRC64;



Query Match 98.9%; Score 1178; DB 1; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-94; Indels 0; Gaps 0;  
 Matches 233; Conservative 0; Mismatches 0;

QY 4 MEIFITGLGASLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLGVDLL 63  
 DB 1 MEIFITGLGASLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLGVDLL 60

QY 64 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAQIIEETPTVPDDTPLGSSAVA 123  
 DB 61 SNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAQIIEETPTVPDDTPLGSSAVA 120

QY 124 TDTNRNRVSVSDKORVWVKPMLAVLTWLPNAYLDAFVFGVGAQYGTGTGRWIFA 163  
 DB 121 TDTNRNRVSVSDKORVWVKPMLAVLTWLPNAYLDAFVFGVGAQYGTGTGRWIFA 180

QY 184 AGAFAASLIWFLVGFCAALSRPLSSPKVWRVNVVAVMTALAKMLMG 236  
 DB 181 AGAFAASLIWFLVGFCAALSRPLSSPKVWRVNVVAVMTALAKMLMG 233

RESULT 2  
 YGGA\_ECOLI  
 ID YGGA\_ECOLI STANDARD; PRT; 211 AA.  
 AC P11667;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein ygga.  
 GN YGGA OR B2923.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blatner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [2]  
 RP SEQUENCE OF 15-211 FROM N.A.  
 RC STRAIN=K12 / CS520;  
 RX MEDLINE=89313302; PubMed=2546007;  
 RA Alefounder P.R., Baldwin S.A., Perham S.A., Short N.J.;  
 RT "Identification, molecular cloning and sequence analysis of a gene  
 RT cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-  
 RT phosphoglycerate kinase and a putative second glyceraldehyde 3-  
 RT phosphate dehydrogenase of Escherichia coli.";  
 RL Mol. Microbiol. 3:723-732 (1989).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).  
 CC -1- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; U28377; AAC69090.1;  
 EMBL; A8000375; AAC75960.1;  
 EMBL; X14436; CA32607.1;  
 PIR; S04736; Q8ECSA.  
 DR EcGene; EGI159; Ygga.  
 DR InterPro; IPR001123; Lyse.  
 DR InterPro; IPR004777; Lys\_exporter.  
 DR Pfam; PF01810; Lyse; 1.

DR TIGRFAME; TIGR00948; 2a75; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 37 57 POTENTIAL.  
 FT TRANSMEM 68 88 POTENTIAL.  
 FT TRANSMEM 111 131 POTENTIAL.  
 FT TRANSMEM 147 167 POTENTIAL.  
 FT TRANSMEM 179 199 POTENTIAL.  
 SQ SEQUENCE 211 AA; 23175 MW; 2DAFE27B6A9B822 CRC64;

Query Match 29.4%; Score 350; DB 1; Length 211;  
 Best Local Similarity 35.7%; Pred. No. 3.6e-23;  
 Matches 81; Conservative 37; Mismatches 75; Indels 34; Gaps 4;

QY 7 FITGLLGASLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLGVDLL 66  
 DB 5 YFQGLGALGAAMTILPLGPQNAFVWQNGIRQYHIMIALCAISDLVLCAGIFGSSALLMQ 64

QY 67 APIVLIDIMRWGGIAYLLWFAVMAAKDAMTKVKAQIIEETPTVPDDTPLGSSAVATDT 126  
 DB 65 SPWLLALVTWGGVAFLLWYFGAFKTMSSNIEL----- 98

QY 127 RNRVRVSVSDKORVWVKPMLAVLTWLPNAYLDAFVFGVGAQYGTGTGRWIFAAG 185  
 DB 99 -----ASAEVMEKQGRW-KIATWLVAVTWNPHVYLDTFVVLGSLGQLDVPEKRW-FALG 151

QY 186 AFAASLIWFLVGFCAALSRPLSSPKVWRVNVVAVMTALAKL 232  
 DB 152 TTSAGFLWFFGLLAAWLAAPRLRTAKAQRINLVVGVGMVFIALQL 198

RESULT 3  
 Y488\_MYCTU  
 ID Y488\_MYCTU STANDARD; PRT; 201 AA.  
 AC Q11154;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV0488.  
 GN RV0488 OR MT0507 OR MTCY2069.14.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekai F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Ruter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence.";  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains.";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; 277162; CAB00949.1; --  
 EMBL; AE006952; AAK44730.1; --  
 TIGR; MT0507; --  
 Tuberculin; RV0488; --  
 InterPro; IPR001123; LysE.  
 InterPro; IPR004777; Lys\_exporter.  
 Pfam; PF01810; LysE; 1.  
 TIGRPFAMs; TIGR00948; 2a75; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 25 45 POTENTIAL.  
 FT TRANSMEM 57 77 POTENTIAL.  
 FT TRANSMEM 104 124 POTENTIAL.  
 FT TRANSMEM 133 153 POTENTIAL.  
 FT TRANSMEM 169 189 POTENTIAL.  
 SQ SEQUENCE 201 AA; 20951 MW; E198975DF088E5E4 CRC64;

Query Match 26.8%; Score 319; DB 1; Length 201;

Best Local Similarity 32.4%; Pred. No. 1.6e-20;  
 Matches 71; Conservative 50; Mismatches 64; Indels 34; Gaps 4;

QY 16 SLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTIGVLLSNAAPVLVDIMR 75  
 DB 3 TLKVAIGPQNAFLVRQIRREYVLVIVLVCGLADGALIAAGVGFGFAALIHAFNPMTLVAR 62  
 QY 76 WGGIAYLLWFVAVMAAKDAMTNKVEAPQIIETEPTVDDTPLGGSVATDTRNRVVEVS 135  
 DB 63 FGGAFLIGYALLAARNW-----RPSGLVPSES--GPAAL----- 96  
 QY 136 VDKQVWVKPMLMAVLVTWLNPNAYLDAFVPIGVGQAQYGTGTGWIFPAAGAFASLIWFP 195  
 DB 97 -----IGVVQMLVTFPLNPHVYLTVDVLIGALANESDL-RWFFGAGAAASVWVFA 148  
 QY 196 LVFGAAALSRPLSSPKVWVWVWVAVVMTALAIAKL 234  
 DB 149 VLGFSGRLQFFPATPAWRILDALVATMIGVAVVVLV 187

# RESULT 4

YJ86\_MYCTU STANDARD; PRT; 199 AA.  
 ID Q10871;  
 AC Q10871;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical protein RV1986  
 GN RV1986 OR MT2040 OR MTCY39.33C.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekalia F.,  
 RA Badcock K., Bigham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence.";  
 RL Nature 393:537-544 (1998).

## RN

RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -|- SIMILARITY: BELONGS TO THE LYS\_E/YGGA FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; 274025; CA98398.1; --

EMBL; AE007056; AAK46315.1; --

TIGR; MT2040; --

Tuberculin; RV1986; --

InterPro; IPR001123; LysE.

InterPro; IPR004777; Lys\_exporter.

Pfam; PF01810; LysE; 1.

TIGRPFAMs; TIGR00948; 2a75; 1.

KW Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 6 26 POTENTIAL.

FT TRANSMEM 42 62 POTENTIAL.

FT TRANSMEM 68 88 POTENTIAL.

FT TRANSMEM 116 136 POTENTIAL.

FT TRANSMEM 144 164 POTENTIAL.

FT TRANSMEM 178 198 POTENTIAL.

SQ SEQUENCE 199 AA; 20775 MW; 6F330132DDCFD0FF CRC64;

Query Match

Best Local Similarity 26.7%; Score 318; DB 1; Length 199;

Matches 70; Conservative 44; Mismatches 27; Indels 34; Gaps 3;

QY 8 ITGLLIGASLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTIGVLLSNA 67

DB 6 VVGFACFTLIAIGAQAFLVLRQIQRHVLVPVALCTVSDIVLIAGIAGFGALIGAH 65

QY 68 PIVLDIMRWGGIAYLLWFVAVMAAKDAMTNKVEAPQIIETEPTVDDTPLGGSVATDTR 127

DB 66 PRALNVVKGGAFLIGYGLLAARRAARFVALIP-----SGATPVR 106

QY 128 NRVREVSVDKORVWVKPMLMAVLVTWLNPNAYLDAFVPIGVGQAQYGTGTGWIFAAAGAF 187

DB 107 -----LAEVLVTCAAFTFLNPHVYLTVDVLIGALANESDQ-RWLFGLGAV 151

QY 188 AASLIWFLVGVGAALSRPLSSPKVWVWVWVAVVMTALAIAKL 232

DB 152 TASAVWFATLGFAGRLGLFTNPGSWRLDGLIIVMMVALGISL 196

## RESULT 5

YGGA\_AERSA

ID YGGA\_AERSA STANDARD; PRT; 206 AA.

AC P70775;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical protein YGGA.

GN YGGA.

OS Aeromonas salmonicida.

OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;

OC Aeromonas.

OX NCBI\_TaxID=645;

```

RN RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 1102;
RX MEDLINE=974311471; PubMed=9286976;
RA Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,
RA Chhabra S.R., Williams P., Macintyre S., Stewart G.S.A.B.;
RT "Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
RT identification of the LuxR homologs AhyRI and AhxRI and their
RT cognate N-acylhomoserine lactone signal molecules.";
RL J. Bacteriol. 179:5271-5281(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U65741; AAB70019.1; ALT_INIT.
CC InterPro; IPR001123; Lyse.
CC InterPro; IPR004777; Lys_exporter.
CC Pfam; PF01810; Lyse; 1.
CC TIGRFAMs; TIGR00948; 2a75; 1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 1 21 POTENTIAL.
CC TRANSMEM 37 57 POTENTIAL.
CC TRANSMEM 65 85 POTENTIAL.
CC TRANSMEM 116 136 POTENTIAL.
CC TRANSMEM 148 168 POTENTIAL.
CC TRANSMEM 185 205 POTENTIAL.
CC SEQUENCE 206 AA; 21505 MW; D1C2C492CDA0179A CRC64;
Query Match 26.0%; Score 310; DB 1; Length 206;
Best Local Similarity 31.9%; Pred. No. 9.5e-20;
Matches 72; Conservative 43; Mismatches 77; Indels 34; Gaps 3;
QY 8 ITGLLGASLLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLVDLLSNAA 67
D 6 LGQFTGLGLAMIIPIGAQNAFVLSRGIHRNHLLATLCCCLDLVLIGVFGGANLAA 65
QY 68 PIVLDIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIEETPTVDDTPLGGSVAATD 127
D 66 PIGLALLTWGVLFCWFGIRSLRSWQGGQ-----GAALADSPR 104
QY 128 NRVREVSVDKQVWVKPML-MAIVLTWLNPNAYLDFAVFIQGVGAQYGDGTGRWIFA 186
D 105 -----LMGVKSVLAWTLGVTLNPNHYLDLMLLGSFGSQFAEPLRPAAGA 152
QY 187 FASLTIWFLVFGGAALSRPLSSPKVWRWVNVVAVMTALAIKL 232
D 153 MLASLVNWFYSLAFGAALSPWLARGVQQAIDTIVGLIMLGLALQL 198
RESULT 6
YGGA_AERHY STANDARD; PRT; 225 AA.
AC P52047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 24.5 kDa protein in ahxR-cdnp intergenic region.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Al;
RX MEDLINE=974311471; PubMed=9286976;
RA Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,

```

```

RA Chhabra S.R., Williams P., Macintyre S., Stewart G.S.A.B.;
RT "Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida:
RT identification of the LuxR homologs AhyRI and AhxRI and their
RT cognate N-acylhomoserine lactone signal molecules.";
RL J. Bacteriol. 179:5271-5281(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X89469; CAA61655.1; ALT_INIT.
CC InterPro; IPR001123; Lyse.
CC InterPro; IPR004777; Lys_exporter.
CC Pfam; PF01810; Lyse; 1.
CC TIGRFAMs; TIGR00948; 2a75; 1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 1 21 POTENTIAL.
CC TRANSMEM 37 57 POTENTIAL.
CC TRANSMEM 65 85 POTENTIAL.
CC TRANSMEM 116 136 POTENTIAL.
CC TRANSMEM 150 170 POTENTIAL.
CC SEQUENCE 225 AA; 24482 MW; 172DB104473B0809 CRC64;
Query Match 21.3%; Score 254; DB 1; Length 225;
Best Local Similarity 26.9%; Pred. No. 6.6e-15;
Matches 61; Conservative 38; Mismatches 74; Indels 54; Gaps 4;
QY 8 ITGLLGASLLLSIGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTGLVDLLSNAA 67
D 6 LGQFTGLGLAMIIPIGAQNAFVLSRGIHRNHLLATLCCCLDLVLIGVFGGANLAA 65
QY 68 PIVLDIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIEETPTVDDTPLGGSVA 123
D 66 PIGLALLTWGVLFCWFGIRSLRSWQGGQAKLADSPQ-----MGYKSV- 111
QY 124 TDTNRVRVSVDRQVWVKPMLMAIVLTWLNPNAYLDFAVFIQGVGAQYGDGTGRWIFA 183
D 112 -----LMTLGVTLNPNHYLDLMLLGSFGSQFAEELRSAPA 149
QY 184 AGAFAASLTIWFLVFGGAALSRPLSSPK-----VWRW 216
D 150 AVAMLASLVNWFYSLAFGAALSPWLARGSQYKSLILLVSPCWGW 196
RESULT 7
RHTC_ECOLI STANDARD; PRT; 206 AA.
ID AC P27846;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine efflux protein.
GN RHTC OR B3823 OR Z5344 OR ECS4753.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RT from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).

```

```

DR Pfam; PF01810; LyseB; 1.
DR TIGRFAMs; TIGR00949; 2A76; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 150 173 POTENTIAL.
SQ SEQUENCE 206 AA; 22474 MW; F64017878CC6D50D CRC64;

Query Match      8.8%; Score 105; DB 1; Length 206;
Best Local Similarity 20.4%; Pred. No. 0.037;
Matches 51; Conservative 36; Mismatches 99; Indels 64; Gaps

Qy 3 IMEIPITGLLGSALLSIGQPQNVLVKQIKR---EGLIAYLLVAVMAAKDMTNK---VEAPQIISETEP 109
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 1 MLMLEFLTVAMHVHVALMSPGPDFFVSQTASVSRKEAMVGVLGTTCGVVMWAGIA-LLG 59
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Qy 60 VDILLSNAPIVLIMRW-----GGIAYLLWFPAVMAAKDMTNK---VEAPQIISETEP 109
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 60 LHL-----IIEKMAWLHTLMVGGGLYLCNMGYQLRGALKKEAVSAPQV----- 106
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Qy 110 TVPDDTFLGGSAVATDTNRNRRVRVEVSVDQRVWVKPMALVLTWLNPNAYLDAPFVPG 169
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 107 -----ELAKGRS-----FLKELLNLANPK-----AIIFYGS 134
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Qy 170 -----VGAQYGDTGRIFAFAGAFAASLIWFPLVFGGAALSRLSPKPVKWRINVVAVV 224
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 135 VFSLFVGDNGVTGARWGIFALIIVETLANFTVVASLFALPOMERGQYRLAKWIDGFAGAL 194
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

Qy 225 MTALAIAKLML 234
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 195 FAGFGIHLLI 204
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||

RESULT 8
YEAS_ECOLI
ID YEAS_ECOLI STANDARD; PRT; 212 AA.
AC P76249; O07971; O07969;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yeas.
GN YEAS OR B1798.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kikagawa M., Kikagawa M., Makino K., Mikami T.,
RA Mizubuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubramaniam S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL RNA SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC CC
```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB000274; AAC74868.1; --  
 DR EMBL; D90823; BAA15593.1; --  
 DR EMBL; D90824; BAA15602.1; --  
 DR EcoGene; EG13505; yeas.  
 DR InterPro; IPR001123; 1.  
 DR Pfam; PF01810; Lyse; 1.  
 DR Hypothetical protein; Transmembrane; Complete proteome.  
 KW TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 12 32 POTENTIAL.  
 FT TRANSMEM 49 69 POTENTIAL.  
 FT TRANSMEM 71 91 POTENTIAL.  
 FT TRANSMEM 122 142 POTENTIAL.  
 FT TRANSMEM 153 173 POTENTIAL.  
 FT TRANSMEM 188 208 POTENTIAL.  
 SQ SEQUENCE 212 AA; 23200 MW; 342E0DF348C9AD9A CRC64;  
 Query Match 8.5%; Score 101; DB 1; Length 212;  
 Best Local Similarity 22.3%; Pred. No. 0.083;  
 Matches 52; Conservative 28; Mismatches 69; Indels 84; Gaps 9;  
 QY 12 LLGASLLLSI-GPQNVILVIRKQIK-REGLIAVLVCLISDFELFIAGTGLVGLDLSNAA 67  
 Db 14 LVGAIFVLVPGNTLFLVKNSSGKGGYLAACGVFIGDAVLMFLAWA-GVATLIKIT 72  
 QY 68 PIVLDNRWGGIAYLLW-----FAVMAAKDAMTNKVEAPQIIEETPTVPDPTLGGSA 121  
 Db 73 PILFNIVRYLGFVLLYGLSKILVATLKGNS-----EAKSDEPQYGAI 116  
 QY 122 VADTNRNRVSVKQVRVWVPMALVILTWLNPNA---YLDFAVFVGGVCAQVGDGTG 178  
 Db 117 -----FKRALISLTNPKAILFYVSFFQFIDVNAHTGIS 152  
 QY 179 RWIFAA-----GAFAA-----SLIWPLVVGFGA 201  
 Db 153 FFILATLELVSPCYLSFLISGAFVQYIRTKKLAKVGNLIIGLWFGFAA 205

RESULT 9  
 ID PTCC ECOLI STANDARD; PRT; 452 AA.  
 AC P17334; P76212; P77332; P76907;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE PTS system, cellobiose-specific IIC component (EIIC-Cel) (Cellobiose-  
 DE permease IIC component) (Phosphotransferase enzyme II, C component).  
 GN CELB OR B1737.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90185127; PubMed=2179047;  
 RA Parker L.L., Hall B.G.;  
 RT "Characterization and nucleotide sequence of the cryptic cel operon  
 RT of Escherichia coli K12.";  
 RL Genetics 124:455-471 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=97251357; PubMed=9097039;  
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,  
 RA Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,  
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,  
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,  
 RA Sampei G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,  
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horuchi T.;  
 RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome  
 RT corresponding to the 28.0-40.1 min region on the linkage map.";  
 RL DNA Res. 3:363-377 (1996).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=91227627; PubMed=2092358;  
 RA Reizer J., Reizer A., Saier M.H. Jr.;  
 RT "The cellobiose permease of Escherichia coli consists of three  
 RT proteins and is homologous to the lactose permease of Staphylococcus  
 RT aureus.";  
 RL Res. Microbiol. 141:1061-1067 (1990).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT  
 CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE  
 CC AND TRANSPORT SYSTEM. THE IIC DOMAINS CONTAIN THE SUGAR BINDING SITE  
 CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY  
 CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS  
 CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO  
 CC THE SUGAR.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -!- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; X52890; CAA37070.1; --  
 DR EMBL; M64438; AAA23547.1; --  
 DR EMBL; AB000268; AAC74807.1; --  
 DR EMBL; D90816; BAA15518.1; --  
 DR EMBL; D90817; BAA15526.1; --  
 DR EMBL; D90818; BAA15532.1; --  
 DR PIR; S10871; S10871.  
 DR EcoGene; EG10141; celB.  
 DR InterPro; IPR004796; Cello\_pts\_IIC.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR InterPro; IPR004501; Pts.  
 DR Pfam; PF02378; PTS\_EIIC\_1.  
 DR TIGRfams; TIGR00359; Cello\_pts\_IIC; 1.  
 DR TIGRfams; TIGR00410; pts; 1.  
 KW Phosphotransferase system; Sugar transport; Transmembrane;  
 KW Inner membrane; Complete proteome.  
 FT TRANSMEM 35 55 POTENTIAL.  
 FT TRANSMEM 86 106 POTENTIAL.  
 FT TRANSMEM 114 134 POTENTIAL.  
 FT TRANSMEM 147 167 POTENTIAL.  
 FT TRANSMEM 187 207 POTENTIAL.  
 FT TRANSMEM 229 249 POTENTIAL.  
 FT TRANSMEM 300 320 POTENTIAL.  
 FT TRANSMEM 344 364 POTENTIAL.  
 FT TRANSMEM 367 387 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT CONFLICT 190 205 LIPGFIILSVGGIIAW -> FNSRLYSFRDGDYCL (IN  
 REF. 1).  
 FT CONFLICT 233 235 VGM -> WL (IN REF. 1).  
 FT CONFLICT 240 265 FVPLWFFGIHGALALDNGIMTP -> LSTALVLRTHA  
 ACADRTGQRHYDA (IN REF. 1).



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AE000140; AAC73431.1; -;  
 DR EMBL; U73857; AAB18053.1; -;  
 DR Ecogen; EG13598; Yahn.  
 DR InterPro; IPR004778; Homoser\_Thr\_eff.  
 DR InterPro; IPR001123; LysE.  
 DR Pfam; PF01810; LysE; 1.  
 DR TIGRFAMs; TIGR00949; 2A76; 1.  
 KW Hypothetical protein; Transmembrane; Complete proteome.  
 FT TRANSMEM 22 42 POTENTIAL.  
 FT TRANSMEM 59 79 POTENTIAL.  
 FT TRANSMEM 85 105 POTENTIAL.  
 FT TRANSMEM 164 184 POTENTIAL.  
 FT TRANSMEM 213 223 POTENTIAL.  
 FT CONFLICT 213 223 LRLIYEGVTQR -> YA (IN REF. 2).  
 SQ SEQUENCE 223 AA; 24811 MW; E17F5ABC31EE3F26 CRC64;

Query Match 7.7%; Score 92; DB 1; Length 223;  
 Best Local Similarity 22.9%; Pred. No. 0.52;  
 Matches 54; Conservative 33; Mismatches 101; Indels 48; Gaps 10;

QY 11 LLLGASLLSISGP-QNVLVIRKQ-----GKREGILAVLLVCLISDVFLFIAGTLGVLLSN 65  
 DB 22 LTVGLFVITPFGANLFFVVVQTSLSARRAGVLTGLVAL-GDAFYSGLGLFGLATLIT 80  
 QY 66 AAPVLDIMRWGGIAYLLWFAVMAAKDAMTKKQAPQIIBETEP-TVPDDTPLGGSAVAT 124  
 DB 81 QCEBIFSLIRIVGGAYLLWFAVMAAKDAMTKKQAPQIIBETEP-TVPDDTPLGGSAVAT 124  
 QY 125 DTRNRVRVSVDRVQVWVKPMLMAVLTWLNPNAYLDADFVFIGV--GAQYDGTGRWIF 182  
 DB 124 -----WYVFFRGLITDLSNPQTUL-FFISFVTLNATPTWARLMA 165  
 QY 183 AAGAFASLIW--PPLVGFGAALSRPLSPKVRWVWVWVMTALAKMLMG 236  
 DB 166 WAGIVLASIIWRVLSQAFSLPAVRRAYG--RMQRVASRVIGAIIGVFALRLIYEG 219

RESULT 12  
 RHTC SALTY STANDARD; PRT; 206 AA.  
 AC Q9L6N7;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Threonine efflux protein.  
 GN RHTC OR STM3959 OR STWD1.31.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OC NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E.; Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Ngwenilik S., Ali J., Dante M., Du P., Hou S., Layman D.,  
 RA Leonard S., Poyoulik S., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan B., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 LT2.";  
 RL Nature 413:852-856(2001).  
 CC -1- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF233324; AAF33433.1; -;  
 DR EMBL; AE008884; AAL22803.1; -;  
 DR StyGene; SG7777; rhtC.  
 DR InterPro; IPR004778; Homoser\_Thr\_eff.  
 DR InterPro; IPR001123; LysE.  
 DR Pfam; PF01810; LysE; 1.  
 DR TIGRFAMs; TIGR00949; 2A76; 1.  
 KW Transport; Transmembrane; Complete proteome.  
 FT TRANSMEM 1 21 POTENTIAL.  
 FT TRANSMEM 44 64 POTENTIAL.  
 FT TRANSMEM 67 87 POTENTIAL.  
 FT TRANSMEM 150 173 POTENTIAL.  
 SQ SEQUENCE 206 AA; 22498 MW; 080477853FC2733F CRC64;

Query Match 7.6%; Score 91; DB 1; Length 206;  
 Best Local Similarity 20.3%; Pred. No. 0.58;  
 Matches 50; Conservative 34; Mismatches 106; Indels 56; Gaps 8;

QY 3 IMEIFTGLGASLLSISGPQNVLVIRKQIKREGILAVLLVCLISDVFLFIAGT--LGV 60  
 DB 1 MWMPLPTVAVHVALMSPDPFFVQSPTAVSRKRAMMVGITCGVMWAGVALLGL 60  
 QY 61 DLLSNAPVLDIMRWGGIAYLLWFAVMAAKDAMTK-----VEAPQIIBETETVPDDTPL 117  
 DB 61 HLIIERKAWLHTIIMVGGGLYLCWGYQMLRGALKKQDAASPHI----- 106  
 QY 118 GGSVAATDTRNRVRVSVDRVQVWVKPMLMAVLTWLNPNAYLDADFVFIGV--VGA 172  
 DB 107 ---ELAQSGRS-----FLKGLLTNLSNPK-----AIYFGSVFSLFVGD 142  
 QY 173 QVGDGTGRWIFAGAF-----ASLIWFFPLVGFGAALSRPLSPKVRWVWVWVMTAL 228  
 DB 143 NVGAAARM---GIFALITLETLANFTVVASLFPALPKMRGRYQRLAKWIDGPGALFAGF 198  
 QY 229 AIKML 234  
 DB 199 GHLLI 204

RESULT 13  
 CHS2\_NEUCR STANDARD; PRT; 944 AA.  
 ID CHS2\_NEUCR  
 AC P30589;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Chitin synthase 2 (EC 2.4.1.16) (Chitin-UDP acetyl-glucosaminyl  
 DE transferase 2).  
 GN CHS-2.  
 OS Neurospora crassa.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 OC NCBI\_TaxID=5141;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95039879; PubMed=7952169;  
 RA Din A.B., Yarden O.;  
 RT "The Neurospora crassa chs-2 gene encodes a non-essential chitin  
 RT synthase.";  
 RL Microbiology 140:2189-2197(1994).  
 RN [2]  
 RP SEQUENCE OF 250-438 FROM N.A.  
 RX MEDLINE=92115692; PubMed=1731323;  
 RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szanislo P.J.,  
 RA Robbins P.W.;







```

FT DOMAIN 1239 1410 WFA 7.
FT DOMAIN 1441 1621 WFA 8.
FT DOMAIN 1641 1814 WFA 9.
FT DOMAIN 1840 2029 WFA 10.
FT DOMAIN 2407 2587 WFA 11.
FT DOMAIN 2625 2821 WFA 12.
FT DOMAIN 2945 3043 FIBRONECTIN TYPE-III.
FT DOMAIN 3068 3137 BPTI/KUNITZ INHIBITOR.
FT SITE 2166 2172 INTERRUPTION IN COLLAGENOUS REGION.
FT SITE 2254 2259 INTERRUPTION IN COLLAGENOUS REGION.
FT SITE 2308 2309 INTERRUPTION IN COLLAGENOUS REGION.
FT SITE 2045 2047 CELL ATTACHMENT SITE.
FT SITE 2153 2155 CELL ATTACHMENT SITE.
FT SITE 2159 2161 CELL ATTACHMENT SITE.
FT ACT_SITE 3082 3083 REACTIVE BOND (BY SIMILARITY).
FT DISULFID 3072 3122 BY SIMILARITY.
FT DISULFID 3081 3105 BY SIMILARITY.
FT DISULFID 3097 3118 BY SIMILARITY.
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2084 2084 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2436 2436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2563 2563 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2581 2581 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2683 2683 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2867 2867 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 3137 AA; 339595 MW; ECB428578B536357 CRC64;

Query Match 7.4%; Score 88.5; DB 1; Length 3137;
Best Local Similarity 25.4%; Pred. No. 18;
Matches 46; Conservative 22; Mismatches 74; Indels 39; Gaps 6;

QY 14 GASLLLSIGPQNVLVKQIK-----REGLIAVLAVCLISDVFLFIAGTIGVLDL----- 63
Db 337 GGSRIEAVPQILVLSGGSSDIRGLAVKQ-----ASIFSISGVNADSAELOQIA 392

QY 64 --SNAAPITVLDIMRWGGIAYLLWFAVMAAKDAMTKNVEAPOIIEETEPTVPDPTPL---G 118
Db 393 TDGSFAFTALDIRNLAARELLLNINIVGVAORLI-LEAPFIVTEVIVNKDILVFLDG 451

QY 119 GSAVATDTRNRVRVEVSVDKQVWVKPMLMAIVLTWLNPNAYLDADFVIGGVGQYGDGTG 178
Db 452 STALGTGPFNSIRDFAKIVQRLVGVGPDLIQAV-----AQYADTV 492

QY 179 R 179
Db 493 R 493

RESULT 15
RHTC_SALTI STANDARD; PRT; 206 AA.
ID_RHTC_SALTI
AC Q8Z3B3;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Threonine efflux protein.
GN RHTC OR STY3600;
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jørgensen K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;

```

```

RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL627278; CAD07933.1; -
DR InterPro; IPR004778; Homoser_Thr_eff.
DR InterPro; IPR001123; Lyse.
DR Pfam; PF01810; Lyse; 1.
DR TIGRfam; TIGR00949; 2A76; 1.
KW Transport; Transmembrane; Complete proteome.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 150 173 POTENTIAL.
SQ SEQUENCE 206 AA; 22480 MW; E9DFCDF4B42D46A0 CRC64;

Query Match 7.4%; Score 88; DB 1; Length 206;
Best Local Similarity 19.9%; Pred. No. 1.1;
Matches 49; Conservative 35; Mismatches 106; Indels 56; Gaps 8;

QY 3 IMEIPITGILLGASLLSIGPQNVLVKQIKREGLIAVLAVCLISDVFLFIAGT---LCV 60
Db 1 MLMLEFTVAMVHVALMSPGDPFFVQSQTAVSRKRKAMMGVLGITCGVMWVAGVALLGL 60

QY 61 DLLSNAAPITVLDIMRWGGIAYLLWFAVMAAKDAMTKN---VEAPQIIEETEPTVPDPTPL 117
Db 61 HLIEKAMWLHTIIVGGGLYLCWGYQMLRGALKKQDAAASPHI----- 106

QY 118 GGSVATDTRNRVRVEVSVDKQVWVKPMLMAIVLTWLNPNAYLDADFVIGG---VGA 172
Db 107 ---ELAQSGRS-----FLKGLLTNLSNPK---AIIYFGSVSLFVGD 142

QY 173 QYGDTCRWIFAGAGAFA-----ASLIWPLVFGGAALSRPLSPKVRWVWVAVVMTAL 228
Db 143 NVGAAARW---GIFALITLETLANFTVVASLFALPKMRGYQRLAKWIDGFAGALFAGF 198

QY 229 AIKML 234
Db 199 GIHLII 204

```

Search completed: April 24, 2003, 18:22:42  
Job time : 13 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: April 24, 2003, 18:21:02 ; Search time 33 Seconds  
(without alignments)  
1473.549 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVMEIFITGLLGASLLLS.....INVVAVVMTALAIKLMIMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	864.5	72.6	231	2 Q8RQM4	Q8RQM4 corynebacte
2	349	29.3	202	16 Q8XXT9	Q8XXT9 ralstonia s
3	347	29.1	211	16 Q8Z3W2	Q8Z3W2 salmonella
4	346	29.1	211	16 Q8ZM68	Q8ZM68 salmonella
5	343	28.8	211	16 Q8XD10	Q8XD10 escherichia
6	336	28.2	212	16 Q8Y2F1	Q8Y2F1 ralstonia s
7	335	28.1	205	16 Q8ZHH6	Q8ZHH6 yersinia pe
8	332	27.9	204	16 Q8HW36	Q8HW36 pseudomonas
9	328.5	27.6	200	16 Q9K4K6	Q9K4K6 streptomyces
10	326.5	27.4	202	16 Q8UGV8	Q8UGV8 agrobacteri
11	302.5	25.4	204	16 Q8Y8G7	Q8Y8G7 brucella me
12	285	23.9	226	16 Q9CKJ7	Q9CKJ7 pasteurella
13	245.5	20.6	211	16 Q9KUN4	Q9KUN4 vibrio chol
14	176.5	14.8	220	16 Q06730	Q06730 bacillus su
15	175.5	14.7	207	16 Q8RHX9	Q8RHX9 fusobacteri
16	173.5	14.6	210	16 Q25420	Q25420 helicobacte

17	170.5	14.3	210	16 Q9ZLC4	Q9ZLC4 helicobacte
18	164	13.8	205	16 Q99V11	Q99V11 staphylococ
19	154.5	13.0	205	16 Q97FY1	Q97FY1 clostridium
20	153.5	12.9	213	16 Q9ZHQ4	Q9ZHQ4 rickettsia
21	151.5	12.7	208	17 Q27538	Q27538 methanobact
22	148	12.4	200	16 Q9KFP7	Q9KFP7 bacillus ha
23	136.5	11.5	206	16 Q8Y2B8	Q8Y2B8 ralstonia s
24	136.5	11.5	213	16 Q9I3A2	Q9I3A2 pseudomonas
25	132	11.1	210	16 Q9AC37	Q9AC37 caulobacter
26	130	10.9	210	16 Q9K8S1	Q9K8S1 bacillus ha
27	129	10.8	210	16 Q983P3	Q983P3 rhizobium l
28	126	10.6	197	16 Q92NS0	Q92NS0 rhizobium m
29	120.5	10.1	208	16 Q9AAY6	Q9AAY6 caulobacter
30	119	10.0	620	16 Q9KMS9	Q9KMS9 vibrio chol
31	118	9.9	234	16 Q8UJW7	Q8UJW7 agrobacteri
32	117.5	9.9	197	16 Q9HZT1	Q9HZT1 pseudomonas
33	116	9.7	205	2 Q9RMX0	Q9RMX0 bacillus an
34	114.5	9.6	208	16 Q9K775	Q9K775 bacillus ha
35	113	9.5	222	16 Q9ABU4	Q9ABU4 agrobacter
36	113	9.5	235	16 Q8UC29	Q8UC29 agrobacteri
37	112.5	9.4	211	16 Q92NX8	Q92NX8 rhizobium m
38	111.5	9.4	208	16 Q98G16	Q98G16 rhizobium l
39	111	9.3	158	2 Q9XBR8	Q9XBR8 zymomonas m
40	111	9.3	210	16 Q8Y176	Q8Y176 ralstonia s
41	111	9.3	212	16 Q8XEY9	Q8XEY9 salmonella
42	109.5	9.2	208	16 Q985J9	Q985J9 rhizobium l
43	109.5	9.2	211	16 Q92LX9	Q92LX9 rhizobium m
44	109.5	9.2	218	16 Q92U06	Q92U06 rhizobium m
45	109.5	9.2	450	17 Q9YC69	Q9YC69 aeropyrum p

## ALIGNMENTS

### RESULT 1

Q8RQM4 PRELIMINARY; PRT; 231 AA.  
AC Q8RQM4;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE\* lysine exporter protein.  
GN LYSE.  
OS Corynebacterium efficiens.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
OC Corynebacterium.  
OX NCBI\_TaxID=152794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Itaya H., Kimura E., Kawahara Y., Sugimoto S.;  
RT "lyse, lyse of Corynebacterium efficiens.";  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB083133; BAB8827.1; -.  
SQ SEQUENCE 231 AA; 24776 MW; 8D14D4D7AD937E17 CRC64;

Query Match 72.6%; Score 864.5; DB 2; Length 231;  
Best Local Similarity 71.2%; Pred. No. 9.3e-65;  
Matches 166; Conservative 30; Mismatches 32; Indels 5; Gaps 2;  
QY 4 MEIFITGLLGASLLLSIGPQNVLVTKQIKREGGLIALLVCLISDVFLFIAGTILGVDLL 63  
Db 4 MEIFVTGLLGASLLLSIAIGPQNVLVTKQIKREGITAVIIIVCLISDVFLFTLGLTGLVGLI 63  
QY 64 SNAAPVILDMRWGGIAYLLWFAVMAAKDAMTKVPAQIIIEETPTVPDDTFLGSAVA 123  
Db 64 SDTAPILDLWCGIAYLLWFAVMAARDALRARTV-TFVEHSEFVAASAGSGG---- 118  
QY 124 TDTNRVRVEVSDQKRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGQAQYGDTCRWIFA 183  
Db 119 VTTKQRPRLRITSGTRQWVRPMLMAIVLTWLNPNAYLDADFVIGVGQAQYGETCRWIFA 178  
QY 184 AGAFAASLLWFFLVGFAAALSRLSPSSPKVWVWVAVVMTALAIKLMIMG 236

[illegible]

Query Match 29.1%; Score 346; DB 16; Length 211;  
 Best Local Similarity 33.6%; Pred. No. 2.2e-21;  
 Matches 79; Conservative 41; Mismatches 81; Indels 34; Gaps 4;

QY 3 IMEIFITGLGASLLISIGPQNVVIVKQIKREGLIAVLVCLISDVFLFIAGTIGVDL 62  
 Db 1 MISYFQGPALGAAMILPLGPQNAFVNOGIRQYHLMIALCALSDVLISAGIFGSA 60  
 QY 63 LSNAAPITLIDIMRWGGIAYLLWFAVMAAKDAMTKVKAPOIIBETEPTVDDTPLGSSAV 122  
 Db 61 LLWQSPWLLALVTWGGVAFLLWTFGFKATMSSNLEL----- 98  
 QY 123 ATDTRNRVRVEVSDQRVWVKPMLMAIVLTWLNPNAYLDADFVIGGVAQYG-DTGRWI 181  
 Db 99 -----ASAEVLKQGRW-KIIATMLAVTWNPHVYLDTFVVLGSLGQLAMEPKRW- 147  
 QY 182 FAAGAFAASLIWPLVGFAGAAALSRPLSSPKVWRWVNVVAVMTALAIKMLMG 236  
 Db 148 FALGTISASFLWFFGLAALAAWLAAPRLRTAKAQRRIINILGVVWMLIAFOLAREG 202

RESULT 5  
 Q8XD10 PRELIMINARY; PRT; 211 AA.  
 AC Q8XD10; (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Orf, hypothetical protein.  
 GN YGGA OR Z4260 OR ECS3794.  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]\_TaxID=83334;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / BDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
 RL Nature 409:529-533 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takani H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli  
 O157:H7 and genomic comparison with a laboratory strain K-12";  
 RL DNA Res. 8:11-22 (2001).  
 DR EMBL; AE005522; AAG58049.1; -  
 DR EMBL; AP002563; BAB37217.1; -  
 DR InterPro; IPR001123; LysE.  
 DR Pfam; PF01810; LysE; 1.  
 DR TIGRFAMs; TIGR00948; 2a75; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 211 AA; 23202 MW; 4422D078B182266C CRC64;

Query Match 28.8%; Score 343; DB 16; Length 211;  
 Best Local Similarity 35.2%; Pred. No. 4e-21;  
 Matches 80; Conservative 37; Mismatches 76; Indels 34; Gaps 4;

QY 7 FITGLLGSALLISIGPQNVVIVKQIKREGLIAVLVCLISDVFLFIAGTIGVDLLSNA 66  
 Db 5 YFOGLAALGAAMILPLGPQNAFVNOGIRQYHLMIALCALSDVLICAGIFGSSALLMQ 64

QY 67 APVLIDIMRWGGIAYLLWFAVMAAKDAMTKVKAPOIIBETEPTVDDTPLGSSAVATDT 126  
 Db 65 SPWLLALVTWGGVAFLLWTFGFKATMSSNLEL----- 98  
 QY 127 RNRVRVEVSDQRVWVKPMLMAIVLTWLNPNAYLDADFVIGGVAQYG-DTGRWIFAAAG 185  
 Db 99 -----ASAEVLKQGRW-KIIATMLAVTWNPHVYLDTFVVLGSLGQLDVEPKRW-FALG 151  
 QY 186 AFAASLIWPLVGFAGAAALSRPLSSPKVWRWVNVVAVMTALAIKL 232  
 Db 152 TISASFLWFFGLAALAAWLAAPRLRTAKSQRRIINLVGVGVWVFIQLQL 198

RESULT 6  
 Q8Y2F1 PRELIMINARY; PRT; 212 AA.  
 AC Q8Y2F1;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Probable transmembrane protein.  
 GN RSC0385 OR RSO3353.  
 OS Ralstonia solanacearum (Pseudomonas solanacearum).  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 OX NCBI\_TaxID=305;  
 RN [1]\_TaxID=305;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GM11000;  
 RX MEDLINE=21681879; PubMed=11823852;  
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
 RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,  
 RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,  
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
 RA Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,  
 RA Weissenbach J., Boucher C.A.;  
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";  
 RL Nature 415:497-502 (2002).  
 DR EMBL; AL646059; CAD13913.1; -  
 DR InterPro; IPR001123; LysE.  
 DR Pfam; PF01810; LysE; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 212 AA; 21896 MW; E29D663C9497EDBB CRC64;

Query Match 28.2%; Score 336; DB 16; Length 212;  
 Best Local Similarity 31.0%; Pred. No. 1.5e-20;  
 Matches 71; Conservative 44; Mismatches 82; Indels 32; Gaps 1;

QY 8 ITGLLGSALLISIGPQNVVIVKQIKREGLIAVLVCLISDVFLFIAGTIGVDLLSNA 67  
 Db 16 LSGFGLASLVAIGAQAQYVLRQGRREYVLRGVLVLCALCDMALIALGVAGMGTLISAH 75  
 QY 68 PIVLDIMRWGGIAYLLWFAVMAAKDAMTKVKAPOIIBETEPTVDDTPLGSSAVATDTR 127  
 Db 76 PAWLTAVERWAGAAFLAYGARAFRAAW-----R 103  
 QY 128 RNRVRVEVSDQRVWVKPMLMAIVLTWLNPNAYLDADFVIGGVAQYDGTGRWIFAAAG 187  
 Db 104 GAERLQARNGDKASHAQVLAASALSLNPHVYLDTVLLGAIGGRYAMPANPAPAGAM 163  
 QY 188 AASLIWPLVGFAGAAALSRPLSSPKVWRWVNVVAVMTALAIKMLMG 236  
 Db 164 CASILWFLSLGFGARLLBPFVAPVAMRVLDALIGAVNMWALTLLMGG 212

RESULT 7  
 Q8ZHH6 PRELIMINARY; PRT; 205 AA.  
 ID Q8ZHH6  
 AC Q8ZHH6;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Putative LysE type translocator.  
GN YPO0918.  
OS Yersinia pestis.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Yersinia.  
RN NCBI\_TaxID=632;  
RP [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;  
RX MEDLINE=21470413; PubMed=11586360;  
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,  
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
RA Feitwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;  
RT "Genome sequence of Yersinia pestis, the causative agent of plague";  
RL Nature 413:523-527(2001).  
DR EMBL; AJ414145; CAC89762.1; -;  
DR InterPro; IPR001123; LysE.  
DR InterPro; IPR004777; Lys\_exporter.  
DR Pfam; PF01810; LysE; 1.  
DR TIGRFAMs; TIGR00948; 2a75; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 205 AA; 22164 MW; 2969524391474CDA CRC64;  
Query Match 28.1%; Score 335; DB 16; Length 205;  
Best Local Similarity 33.1%; Pred. No. 1.8e-20;  
Matches 80; Conservative 44; Mismatches 70; Indels 48; Gaps 6;  
QY 3 IMEFTFGLGASLLSISGPONVIVIKQIKREGIIVLLVCLISDVFLFIAGTIGVDL 62  
DB 1 MLAVYLHGFIISAAILPLGPQNVFNQGIKQKHLMSSALCALSDIILICAGIFGSA 60  
QY 63 LSNAAPIVLDIMRWGGIAYLLWF---AVMAAKDAMTKNVEAPQIIETEPTVDDTPLGG 119  
DB 61 LLSRSPILLALVWGGVAFPLMWYWGALMAA-----WRG 94  
QY 120 SAVATD---TRNRVRVEVSDQRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGQAQ-Y 174  
DB 95 DGVASSATSVTQGRWILVTL-----LAVTLNPHVYLDTFVVLGSLGQLL 141  
QY 175 GDTGRWIFAFAGFAASLPLVGVFGAAALSRPLSSPKVWRWVNVVAVMTALAKML 234  
DB 142 PDIRPW-FALGAVTASIVWFFALALLAALWLSPLNRPVQAQRIINLFGVGMGFIAPQLAR 200  
QY 235 MG 236  
DB 201 QG 202  
RESULT 8  
Q9HW36 PRELIMINARY; PRT; 200 AA.  
AC Q9HW36  
DT 01-MAR-2001 (TEMBLrel. 16, Created)  
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)  
DT 01-OCT-2001 (TEMBLrel. 18, Last annotation update)  
DE Probable transporter.  
GN PA4365.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OX Pseudomonas.  
RN NCBI\_TaxID=287;  
RP [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PAO1;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey R.K., Brinkman P.S.L., Huftnagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an  
RT opportunistic pathogen";  
RL Nature 406:959-964(2000).  
DR EMBL; AE004852; AAG07753.1; -;  
DR InterPro; IPR002109; Glutaredoxin.  
DR InterPro; IPR001123; LysE.  
DR Pfam; PF01810; LysE; 1.  
DR PROSITE; PS00195; GLUTAREDOXIN; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 200 AA; 21144 MW; 7C10CB8D8FE36A4 CRC64;  
Query Match 27.9%; Score 332; DB 16; Length 200;  
Best Local Similarity 33.5%; Pred. No. 3.1e-20;  
Matches 77; Conservative 42; Mismatches 77; Indels 34; Gaps 4;  
QY 7 FITGLLGASLLSISGPONVIVIKQIKREGIIVLLVCLISDVFLFIAGTIGVDLSNA 66  
DB 5 YLNGILVAAGLIIAIGAQNAFVLAQSLRREHLSVAALCVFCDAVLVSLGVFLAKLLE 64  
QY 67 APIVLDIMRWGGIAYLLWFVAVMAAKDAMTKNVEAPQIIETEPTVDDTPLGGSAVATDT 126  
DB 65 NPTLLAIARWGGIAYLLTWYGLKALLALKR-----PDAL---GNAETGP 105  
QY 127 RNRVRVEVSDQRVWVKPMLMAIVLTWLNPNAYLDADFVIGVGQAQYDGTGRWIFAAGA 186  
DB 106 RSRKAV-----LLAALAVTLNPHVYLDTFVVLGSLGQAQQAAPG--AYALGA 150  
QY 187 FASLITWPLVFGAAALSRPLSSPKVWRWVNVVAVMTALAKMLMG 236  
DB 151 ASASLWFFALGALAAWLANPWLARPATWRLDLWVAAMLMGAAOULLFRG 200  
RESULT 9  
Q9K4K6 PRELIMINARY; PRT; 204 AA.  
AC Q9K4K6  
DT 01-OCT-2000 (TEMBLrel. 15, Created)  
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
DE Putative membrane transport protein.  
GN SC07308 OR SC5F8.18.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Seeger K.J., Harris D.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RP [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RP [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8843436;  
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome";  
RL Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,



RA	Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.
RA	McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA	Fraser C.M.;
RT	"DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT	cholerae.";
RL	Nature 406;477-483(2000).
DR	EMBL; AE004134; AAF93654.1; -.
DR	TIGR; VC0481; -.
DR	InterPro; IPR001123; LySE.
DR	Pfam; PF01810; LySE; 1.
KW	Complete proteome.
SQ	SEQUENCE 211 AA; C92B752CCEGB0741 CRC64;
Query Match	20.6%; Score 245.5; DB 16; Length 211;
Best Local Similarity	28.6%; Pred. No. 5.8e-13;
Matches	Conservative 34; Mismatches 97; Indels 31; Gaps
QY	6 IFITGLLGGASLLLSIGFONVVIKQIKREGHIAVLVCLISDVFLFIAGTILGVDILLN 65
Dd	:     :     :     :     :     :     :     :     :     :   :
Dd	5 ILQGFSLGATMIPIGAQNAYVLNQIKRHHLHTAATCGVLDIMFITILGIFGGALIS 64
QY*	66 AAPVILDIMRWGGIAYLLFAVMAAKDAMTKNEAPQIIIEETPTVPDDTLFGSSAVTD 125
Dd	:     :     :     :     :     :     :     :     :     :   :
Dd	65 QNTSLLIIGTVTAGLIFLCGGYSLSRAL----KKPQSESTA-----NPMAAGRKA- 113
QY	126 TRNRVRVEVSVDQRVMVKMPLMAVLTLNPNLAYLDAFVFIVGVGAQYGDTGRWFPAAG 185
Dd	:     :     :     :     :     :     :     :     :     :   :
Dd	114 -----IFGAFATVNPHLYLTDTVILSIGSGQFOGDERISPAG 153
QY	186 AFRASLIWFFLVPGGAALSRLSPKVWRWINVVAVVMTALAIDL 232
Dd	:     :     :     :     :     :     :     :     :     :   :
Dd	154 TILASFVWFFTLSIGAALKSTLTSLKPVRVIDMAVAAMFIIAIAL 200
RESULT 14	
O06730	PRELIMINARY; PRT; 220 AA.
ID	O06730
AC	O06730;
DT	01-JUL-1997 (TEMBLrel. 04, Created)
DT	01-JAN-1998 (TEMBLrel. 05, last sequence update)
DT	01-JUN-2002 (TEMBLrel. 21, last annotation update)
DE	YISU protein.
GN	YISU.
OS	Bacillus subtilis.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC	Bacillaceae; Bacillus.
RX	NCBI_TaxID=1423;
OX	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=98015416; PubMed=9353932;
RA	Roche B., Autret S., Levine A., Vannier F., Medina N., Seror S.J.;
RT	"A Bacillus subtilis chromosome segment at the 100 degrees to 102
RT	degrees position encoding 11 membrane proteins.";
RL	Microbiology 143:3309-3312(1997).
[2]	
RC	SEQUENCE FROM N.A.
PC	STRAIN=168;
RX	MEDLINE=98044033; Pubmed=9384377;
RA	Aune F., Ogawara N., Moszer I., Albertini A.M., Alloni G.,
RA	Arvedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherst S.,
RA*	Bouriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA	Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA	Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA	Ghim S.Y., Glauber P., Goiffau A., Golightly E.J., Grandi G.,
RA	Guiseppi G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,
RA	Jalilbert H., Holsappel S., Hosono S., Hulio M.P., Itaya M., Jones L.,
RA	Harris B., Karanata D., Kasahara Y., Klaerr-Bleichard M., Klein C.,
RA	Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA	Kurita K., Lapidos A., Lardinotis S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Muel C., Medigue C.,  
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,  
 RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weizensegger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the gram-positive bacterium Bacillus  
 RT subtilis.";  
 RL Nature 390:249-256 (1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Y09476; CAA70650.1; -;  
 DR EMBL; Z99109; CAB12926.1; -;  
 DR InterPro; IPR001123; Lyse.  
 DR InterPro; IPR004777; Lys\_exporter.  
 DR Pfam; PF01810; LyseE; 1.  
 DR TIGRFAMs; TIGR00948; 2a75; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 220 AA; 24065 MW; C18C2810F44F9656 CRC64;

Query Match 14.8%; Score 176.5; DB 16; Length 220;  
 Best Local Similarity 22.9%; Pred. No. 3.6e-07;  
 Matches 53; Conservative 36; Mismatches 99; Indels 43; Gaps 4;  
 QY 4 MEIFITGLLGASLLSIGPQNVLVKQIGKREGL---TAVLVCLISDVFLFIAGTLGV 60  
 DB 18 MNAIHGIVLAFGLILPLGVQNVFIQOGALOKHWRALPAVISASVCDTLILVLAVAGV 77  
 QY 61 DLLSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGGS 120  
 DB 78 SVIVQELPVPETVMAGGFLFLYMGV-----105  
 QY 121 AVATDTRNRVRVSVVDKQVWV--RPLMAI VLTWLNPNAYLDFAVFIGGVGAQYGDG 178  
 DB 106 -----TWN-IRPNTSQNEKHTFPKQAAFAAAVSLNPHAILDTIGVIGTSLSQSGLE 159  
 QY 179 RMTFAGAGFAASLIWFLVGFAGAAALSRPLSSPKVWRWNVVAVVMTALA 229  
 DB 160 KWLFAACIAVSWIWFISLAIGRLFTDTSGRLMLIVNKCSAAVWMAAA 210

RESULT 15  
 Q8RH9 PRELIMINARY; PRT; 207 AA.  
 AC Q8RH9;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE L-lysine permease.  
 GN FN1861.  
 OS Fusobacterium nucleatum (subsp. nucleatum).  
 OC Bacteria; Fusobacteria; Fusobacterium.  
 OX NCBI\_TaxID=76856;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 25586;  
 RX MEDLINE=21886394; PubMed=11889109;  
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,  
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltzman E., Bernal A.,  
 RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
 Fonstein M., Kyrpides N., Overbeek R.;

RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
 RT nucleatum strain ATCC 25586.";  
 RL J. Bacteriol. 184:2005-2018 (2002).  
 DR EMBL; AE010488; AAL93960.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 207 AA; 23192 MW; 5833132210F1A771 CRC64;  
 Query Match 14.7%; Score 175.5; DB 16; Length 207;  
 Best Local Similarity 24.5%; Pred. No. 4.1e-07;  
 Matches 58; Conservative 35; Mismatches 97; Indels 47; Gaps 5;  
 QY 4 MEIFITGLLGASLLSIGPQNVLVKQIGKREGIAYLVCLISDVFLFIAGTLGYD 61  
 DB 1 MEKYLQGLMGLAYVAPIGVQNLVINSAITQKRSKALLIALIVIFEDVTLAFACFFGIG 60  
 QY 62 DLLSNAAPVLDIMRWGGIAYLLWFAVMAAKDAMTKNVEAPQIIETEPTVPDDTPLGSSA 121  
 DB 61 LL-----IDKLEWLKLIIL-----VGSIV 80  
 QY 122 VATDTRNRVRVSVVDKQVWVVKPMLMAI---VLTWLNPNAYLDFAVFIGGVGAQYGD 177  
 DB 81 IYIGGGLRSKSELKNDNDMDIPLLKAITSCVVTWFPQAIIDGTWMLGAFRATLSSE 140  
 QY 178 GRWIFAGAGFAASLIWFLVGFAGAAALSRPLSSPKVWRWNVVAVVMTALA 234  
 DB 141 AGIYFPLGVTSASFQWFMGLSIFISLFSHKFND-KVLKVINVCGLVIIFYGVKLL 196

Search completed: April 24, 2003, 18:23:23  
 Job time : 36 secs





GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:20:02 ; Search time 36 Seconds  
(without alignments)  
873.532 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVIMEIFITCLLGASLLLS.....INVVAVVMTALAKLMLMG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1191	100.0	236	18	AAW37715
2	1191	100.0	236	22	AAW37715
3	1178	98.9	233	22	AAU71888
4	1178	98.9	233	22	AAU71888
5	864.5	72.6	231	22	AAU79660
6	350	29.4	211	21	AAU79660
7	170.5	14.3	210	18	AAU71888
8	170.5	14.3	215	18	AAU71888
9	132.5	11.1	226	22	AAU71888
10	119.5	10.0	173	23	ABP39181
					C. glutamicum Lys
					C. glutamicum prote
					C. glutamicum meta
					Corynebacterium gl
					Corynebacterium th
					Escherichia coli Y
					H. pylori surface
					H. pylori transmem
					C glutamicum prote
					Staphylococcus epi

11	114.5	9.6	229	21	AAU11637	A. vitis hypersens
12	105	8.8	206	21	AAU99598	E. coli L-threonin
13	101	8.5	212	21	AAU01787	Escherichia coli Y
14	95	8.0	390	22	AAU58229	Propionibacterium
15	93.5	7.9	452	22	AAU34590	E. coli cellular p
16	93	7.8	459	22	AAU93244	C glutamicum prote
17	93	7.8	459	22	AAU78987	C. glutamicum SRT
18	92	7.7	223	21	AAU01786	Escherichia coli Y
19	92	7.7	694	23	ABU93711	Herbicidally activ
20	91.5	7.7	446	22	AAU92602	C glutamicum prote
21	91.5	7.7	446	22	AAU76722	Corynebacterium gl
22	91.5	7.7	446	23	AAU22703	Corynebacterium gl
23	91	7.6	111	20	AAU89961	Antigen from clust
24	91	7.6	624	22	AAU90535	C glutamicum prote
25	91	7.6	628	22	AAU76784	Corynebacterium gl
26	91	7.6	628	22	AAU76784	Corynebacterium gl
27	90	7.6	4572	19	AAU52845	A. mediterranei ri
28	89.5	7.5	205	21	AAU79298	E. coli L-homoseri
29	89.5	7.5	205	21	AAU79298	E. coli RhtB prote
30	89.5	7.5	340	22	AAU36387	Pseudomonas aerugi
31	89	7.5	123	22	ABG23207	Novel human diagno
32	89	7.5	716	22	AAU64250	Propionibacterium
33	88	7.4	406	22	AAU39609	Propionibacterium
34	88	7.4	1308	22	ABU62870	Drosophila melanog
35	86.5	7.3	195	21	AAU01788	Escherichia coli Y
36	86.5	7.3	279	22	AAU55142	Propionibacterium
37	86	7.2	295	22	AAU34511	E. coli cellular p
38	86	7.2	400	22	AAU90734	C glutamicum prote
39	85.5	7.2	138	22	AAU82080	S. epidermidis ope
40	85.5	7.2	224	23	ABP39763	Staphylococcus epi
41	85	7.1	548	22	ABU62090	Drosophila melanog
42	84.5	7.1	415	23	AAU22164	Ramoplanin biosynt
43	84.5	7.1	810	23	ABU93675	Herbicidally activ
44	84	7.1	514	23	ABU93675	Herbicidally activ
45	83.5	7.0	210	22	AAU81807	S. epidermidis ope

#### ALIGNMENTS

RESULT 1  
AAW37715  
ID AAW37715 standard; Protein; 236 AA.  
XX  
AC AAW37715;  
XX  
DT 12-MAR-1998 (first entry)  
XX  
DE C. glutamicum Lys E protein (lysine export protein).  
XX  
KW LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;  
KW Microbial production; amino acid; animal feed additive.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN DE19548222-A1.  
XX  
PD 26-JUN-1997.  
XX  
PF 22-DEC-1995; 95DE-1048222.  
XX  
PR 22-DEC-1995; 95DE-1048222.  
XX  
(KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
XX  
PA Eggeling L, Sahn H, Vrljic M;  
PI  
XX  
DR WPI; 1997-333867/31.  
DR N-PSDB; AAT96816.  
XX  
PT Increasing microbial production of amino acids, especially lysine -  
PT by improving export carrier activity or corresponding gene  
PT expression, also new export and regulatory genes from

PT Corynebacterium  
XX  
PS Claim 42; Page 10; 16pp; German.  
XX

This sequence is the LysE protein product, a lysine export protein.  
CC LysG and LysE encode a lysine transport regulatory protein and an export  
CC protein, respectively. Microbial production of amino acids (A) is  
CC improved by increasing the export-carrier activity and/or the export gene  
CC expression in a microorganism that produces (A). The method is  
CC specifically used to increase production of lysine, used as an animal  
CC feed additive. Other (A) are variously useful as pharmaceuticals,  
CC condiments and intermediates for fine chemicals. This method increases  
CC the amount of (A) secreted into the culture medium. Export of (A) has  
CC been found to depend on a single gene.  
XX

SQ Sequence 236 AA;  
Query Match 100.0%; Score 1191; DB 18; Length 236;  
Best Local Similarity 100.0%; Pred. No. 2e-128;  
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVIMEIFITGLLGASLLSISGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60  
DB 1 MVIMEIFITGLLGASLLSISGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60  
QY 61 DLLSNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETPTVDDTPLGGS 120  
DB 61 DLLSNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETPTVDDTPLGGS 120  
QY 121 AVATDTRNRVRVEVSVDKQVWVKPMLMAIVLTWLNPNAYLDADFVIGGVGAQYGDTRW 180  
DB 121 AVATDTRNRVRVEVSVDKQVWVKPMLMAIVLTWLNPNAYLDADFVIGGVGAQYGDTRW 180  
QY 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236  
DB 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236

RESULT 2  
AAG93201  
ID AAG93201 standard; Protein; 236 AA.  
XX  
AC AAG93201;  
XX  
DT 26-SEP-2001 (first entry)  
XX  
DE C glutamicum protein fragment SEQ. ID NO: 6955.  
XX  
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis.  
XX  
OS Corynebacterium glutamicum.  
XX  
XX  
XX EPI108790-A2.  
XX  
XX 20-JUN-2001.  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
XX  
XX 16-DEC-1999; 99JP-0377484.  
XX 07-APR-2000; 2000JP-0159162.  
XX 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
XX Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI; 2001-376931/40.  
XX N<sub>2</sub>PSDB; AAH68420.  
XX  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
XX mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
PS Claim 29; SEQ ID NO: 6955; 246pp + Sequence Listing; English.  
XX

The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or identifying a homologue of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX

SQ Sequence 236 AA;  
Query Match 100.0%; Score 1191; DB 22; Length 236;  
Best Local Similarity 100.0%; Pred. No. 2e-128;  
Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVIMEIFITGLLGASLLSISGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60  
DB 1 MVIMEIFITGLLGASLLSISGPQNVLVKQIKREGLIAVLLVCLISDVFLFIAGTLGV 60  
QY 61 DLLSNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETPTVDDTPLGGS 120  
DB 61 DLLSNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETPTVDDTPLGGS 120  
QY 121 AVATDTRNRVRVEVSVDKQVWVKPMLMAIVLTWLNPNAYLDADFVIGGVGAQYGDTRW 180  
DB 121 AVATDTRNRVRVEVSVDKQVWVKPMLMAIVLTWLNPNAYLDADFVIGGVGAQYGDTRW 180  
QY 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236  
DB 181 IFAAGAFASLIWFLVGFAGAAALSRPLSSPKVWRWVNVVAVVMTALAIKMLMG 236

RESULT 3  
AAU71888  
ID AAU71888 standard; Protein; 233 AA.  
XX  
AC AAU71888;  
XX  
DT 26-FEB-2002 (first entry)  
XX  
DE C. glutamicum metabolic pathway protein encoded by gene #23.  
XX  
KW Metabolic pathway protein; MP; lysine biosynthesis pathway;  
XX methionine biosynthesis pathway; large-scale production of fine chemical;  
XX Corynebacterium diphtheriae; diphtheria.  
XX  
XX Corynebacterium glutamicum.  
XX  
XX WO200166573-A2.  
XX  
XX 13-SEP-2001.  
XX  
XX 22-DEC-2000; 2000WO-IB02035.  
XX  
XX 09-MAR-2000; 2000US-187970P.  
XX 23-JUN-2000; 2000US-0606740.  
XX  
XX (BADI ) BASF AG.  
XX  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;  
XX Lee H, Hwang B;  
XX WPI; 2001-582269/65.  
XX

DR N-PSDB; AAS96098.  
 XX  
 PT Nucleic acids encoding metabolic pathway proteins from Corynebacterium  
 PT glutamicum, useful for producing methionine and lysine in  
 PT Corynebacterium and Brevibacterium -  
 XX  
 PS Disclosure; Page 215-216; 316pp; English.  
 XX  
 CC The present invention relates to the isolation of novel Corynebacterium  
 CC glutamicum genes (AAS96073-AAS96132) encoding metabolic pathway (MP)  
 CC proteins. The metabolic pathway proteins of the invention  
 CC include enzymes involved in the lysine and methionine biosynthetic  
 CC pathways. The polynucleotide sequences of the invention can be used  
 CC for the large-scale production and/or modulation of expression of the  
 CC fine chemicals such as lysine and methionine. The sequences of the  
 CC invention may be used to identify C. glutamicum and related organisms  
 CC e.g. C. diphtheriae in a subject to detect diphtheria.  
 CC AAU71863-AAU71922 represent the novel C. glutamicum metabolic pathway  
 CC proteins of the invention.  
 XX  
 SQ Sequence 233 AA;  
 Query Match 98.9%; Score 1178; DB 22; Length 233;  
 Best Local Similarity 100.0%; Pred. No. 6.1e-127;  
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 MEIFITGLLGASLLLSIGFQNVIVIKQIKREGILAVLVCLISDVFLFTAGTGLVDLL 63  
 DB 1 MEIFITGLLGASLLLSIGFQNVIVIKQIKREGILAVLVCLISDVFLFTAGTGLVDLL 60  
 QY 64 SNAAPVLDIMRWGGIAYLWFAVMAAKDAMTKVEAPQIIETEPTVPDTPLGSSAVA 123  
 DB 61 SNAAPVLDIMRWGGIAYLWFAVMAAKDAMTKVEAPQIIETEPTVPDTPLGSSAVA 120  
 QY 124 TDTNRVRVSVVDKQVWVKPMLMAVLVTWLNPNAYLDFAVFVGGVGAQYGTGRWIFA 183  
 DB 121 TDTNRVRVSVVDKQVWVKPMLMAVLVTWLNPNAYLDFAVFVGGVGAQYGTGRWIFA 180  
 QY 184 AGAFAASLIWFLVGFAGAAALSRPLSPKVRWVNVVAVVMTALAKLMLMG 236  
 DB 181 AGAFAASLIWFLVGFAGAAALSRPLSPKVRWVNVVAVVMTALAKLMLMG 233  
 RESULT 4  
 AAB79660  
 ID AAB79660 standard; Protein; 233 AA.  
 AC AAB79660;  
 XX  
 XX 30-APR-2001 (first entry)  
 DT  
 DE Corynebacterium glutamicum MP protein sequence SEQ ID NO:54.  
 XX  
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;  
 KW fine chemical production; microorganism; organic acid; nucleoside;  
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;  
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;  
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100843-A2.  
 XX  
 PD 04-JAN-2001.  
 XX  
 PF 23-JUN-2000; 2000WO-IB00923.  
 XX  
 PR 25-JUN-1999; 99US-0141031.  
 PR 01-JUL-1999; 99DE-1030476.  
 PR 02-JUL-1999; 99US-0142101.  
 PR 08-JUL-1999; 99DE-1031415.  
 PR 08-JUL-1999; 99DE-1031418.  
 PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.  
 PR 08-JUL-1999; 99DE-1031424.  
 PR 08-JUL-1999; 99DE-1031428.  
 PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031435.  
 PR 08-JUL-1999; 99DE-1031443.  
 PR 08-JUL-1999; 99DE-1031453.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031465.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 08-JUL-1999; 99DE-1031573.  
 PR 08-JUL-1999; 99DE-1031592.  
 PR 08-JUL-1999; 99DE-1031632.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 09-JUL-1999; 99DE-1032125.  
 PR 09-JUL-1999; 99DE-1032126.  
 PR 09-JUL-1999; 99DE-1032130.  
 PR 09-JUL-1999; 99DE-1032186.  
 PR 09-JUL-1999; 99DE-1032206.  
 PR 09-JUL-1999; 99DE-1032227.  
 PR 09-JUL-1999; 99DE-1032228.  
 PR 09-JUL-1999; 99DE-1032229.  
 PR 09-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032232.  
 PR 14-JUL-1999; 99DE-1032926.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1033004.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 12-AUG-1999; 99US-0148613.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041380.  
 PR 31-AUG-1999; 99DE-1041394.  
 PR 31-AUG-1999; 99DE-1041396.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042124.  
 PR 03-SEP-1999; 99DE-1042129.  
 PR 09-MAR-2000; 2000US-0187970.

(BADI ) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

WPI; 2001-137957/14.

N-PSDB; AAF71779.

Nucleic acids from Corynebacterium glutamicum encoding metabolic  
 pathway proteins, useful for producing fine chemicals in  
 microorganisms, including organic acids, nonproteinogenic amino acids,  
 and purine and pyrimidine bases -

Claim 20; Page 234-235; 1737pp; English.

AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic  
 pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum  
 MP nucleic acids are useful for the production of fine chemicals  
 in microorganisms, including organic acids, nonproteinogenic amino  
 acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
 saturated and unsaturated fatty acids, diols, carbohydrates, aromatic

CC compounds, vitamins, cofactors, polyketides and enzymes.

XX  
SQ Sequence 233 AA;  
Query Match 98.9%; Score 1178; DB 22; Length 233;  
Best Local Similarity 100.0%; Pred. No. 6.1e-127;  
Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 63  
DB 1 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 60  
QY 64 SNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 123  
DB 61 SNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 120  
QY 124 TDTNRVRVSVVDKQVRWVKPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 183  
DB 121 TDTNRVRVSVVDKQVRWVKPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 180  
QY 184 AGAPAAASLIWFPPLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKMLMG 236  
DB 181 AGAPAAASLIWFPPLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKMLMG 233  
RESULT 5  
AAG64047  
ID AAG64047 standard; Protein; 231 AA.  
AC AAG64047;  
XX  
DT 11-SEP-2001 (first entry)  
XX  
DE Corynebacterium thermoaminogenes lysin biosynthetic enzyme lyse.  
XX  
KW Heat-resistant; lysin biosynthesis; enzyme; coryneform;  
KW aspartate-semialdehyde dehydrogenase; lyse.  
XX  
OS Corynebacterium thermoaminogenes.  
XX  
PN JP2001120270-A.  
XX  
PD 08-MAY-2001.  
XX  
PF 01-NOV-1999; 99JP-0311148.  
XX  
PR 01-NOV-1999; 99JP-0311148.  
XX  
PA (AJIN ) AJINOMOTO KK.  
XX  
DR WPI; 2001-364760/38.  
DR N-PSDB; AAH45375.  
XX  
XX  
PT A heat-resistant lysin biosynthetic system enzyme gene of a high  
PT temperature-resistant coryneform microbe -  
PS Claim 7; Page 24; 27pp; Japanese.  
XX  
CC The invention relates to a gene from a high temperature-resistant  
CC coryneform microbe that encodes a heat-resistant lysin biosynthetic  
CC enzyme. The enzyme has aspartate-semialdehyde dehydrogenase activity  
CC and can be used for growing amino acid-producing microbes. The  
CC present amino acid sequence corresponds to an enzyme of the  
CC invention.  
XX  
SQ Sequence 231 AA;  
Query Match 72.68%; Score 864.5; DB 22; Length 231;  
Best Local Similarity 71.24%; Pred. No. 6.7e-91;  
Matches 166; Conservative 30; Mismatches 32; Indels 5; Gaps 2;  
QY 4 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 63  
DB 1 MEIPITGLLGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVDLL 60

DB 4 MEIPITGLLGASLLLSIGPQNVLVKQIKREGITAVIIVCLISDVFLFIAGTGLGVLLI 63  
QY 64 SNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVA 123  
DB 64 SDTAPILIDILRWCGIAYLLWFAVMAARDALRARTEV-TFVEHSEPVAAASASGGG---- 118  
QY 124 TDTNRVRVSVVDKQVRWVKPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 183  
DB 119 VTTKQRPRLRITSGTQVWVRPMLMAIVLTWLNPNAYLDFAVFVGGVGAQYGDTRWIFA 178  
QY 184 AGAPAAASLIWFPPLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKMLMG 236  
DB 179 AGAPAAASLIWFPPLVGFCAALSRPLSSPKVWRWVNVVAVMTALAIAKMLMG 231  
RESULT 6  
AAB01789  
ID AAB01789 standard; Protein; 211 AA.  
XX  
AC AAB01789;  
XX  
DT 03-JAN-2001 (first entry)  
XX  
DE Escherichia coli YggA amino acid excretion protein.  
XX  
KW E. coli; YggA gene; amino acid production; excretion protein gene;  
KW amino acid excretion protein.  
XX  
OS Escherichia coli.  
XX  
PN EP1016710-A2.  
XX  
PD 05-JUL-2000.  
XX  
PF 17-DEC-1999; 99EP-0125263.  
XX  
PR 30-DEC-1998; 98RU-0124016.  
PR 09-MAR-1999; 98RU-0104431.  
XX  
PA (AJIN ) AJINOMOTO CO INC.  
XX  
PI Livshits VA, Zakataeva NP, Nakanishi K, Alekhin VV, Troshin PV;  
PI Tokhmakova IL;  
XX  
DR WPI; 2000-414802/36.  
DR N-PSDB; AAA52691.  
XX  
PT Increased production of L-amino acids by an Escherichia bacterium  
PT comprises increasing the expression amount of an L-amino acid excretion  
PT protein -  
XX  
PS Claim 1; Page 25; 29pp; English.  
XX  
CC The present sequence is the YggA amino acid excretion protein from  
CC Escherichia coli. This protein is involved in the production of amino  
CC acids, and an increase in its expression leads to an increased  
CC accumulation of amino acids in the cell. In this case, an increase in  
CC arginine, glutamic acid and lysine is achieved if multiple copies of the  
CC gene are transfected into a bacterium. The bacterium used is E. coli.  
XX  
SQ Sequence 211 AA;  
Query Match 29.4%; Score 350; DB 21; Length 211;  
Best Local Similarity 35.7%; Pred. No. 8.4e-32;  
Matches 81; Conservative 37; Mismatches 75; Indels 34; Gaps 4;  
QY 7 FITGALGASLLLSIGPQNVLVKQIKREGIAVLVCLISDVFLFIAGTGLGVLLSNA 66  
DB 5 YFOGLAGAAITLPLGPFQNAFVNMNQIRYHIMTIALCAISDLVLCAGIFGGSALLMQ 64  
QY 67 APTVLDIRWGGIAYLLWFAVMAAKDAMTNKVEAPOIIEETEPTVPDDTPLGGSVAATDT 126  
DB 65 SPWLLALVTWGGVAFLLWYGFCAFTAMSSNIEL----- 98

QY 127 RNRVRVEVSDKQVWVKPMLMAVLTLNPNAYLDAFVFIGGVAQVG-DTGRWIFAAG 185  
 Db 99 -----ASAENVKQGRW-KIATMLAVTWNPHVYLDTFVVLGSLGGQLDVEPKW-FALG 151  
 QY 186 APAASLIWFPLVGGAAALSRPLSSPKVRWVWVAVVMTALAIAKL 232  
 Db 152 TISASFLWFFGLALLAAWLAFLRLRTAKAQRITINLVGCVWVFIALQL 198

RESULT 7  
 AAW20426  
 ID AAW20426 standard; Protein; 210 AA.  
 XX AAW20426;  
 XX  
 DT 14-JUL-1997 (first entry)  
 XX  
 DE H. pylori surface membrane protein 33986087.aa.  
 XX  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO9640893-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 06-JUN-1996; 96WO-US09122.  
 XX  
 PR 01-APR-1996; 96US-0630405.  
 PR 07-JUN-1995; 95US-0487032.  
 XX  
 PA (ASTR ) ASTRA AB.  
 XX  
 PI Berglindh OT, Smith D, Mellgaard BL;  
 XX WPI: 1997-052306/05.  
 DR N-PSDB; AAT67601.  
 DR  
 XX Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 XX  
 PS Claim 73; Page 604; 1481pp; English.  
 XX  
 CC The present sequence shows a Helicobacter pylori surface membrane  
 CC protein, having five transmembrane regions.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.  
 XX  
 SQ Sequence 210 AA;  
 Query Match 14.3%; Score 170.5; DB 18; Length 210;  
 Best Local Similarity 21.5%; Pred. No. 3.6e-11;  
 Matches 50; Conservative 48; Mismatches 92; Indels 43; Gaps 5;

QY 6 IFITGLLGSALLSTGPNVIVIKQIKREGIIVLVCLISDVFLTAGTLGVLLSN 65  
 Db 4 VFIEGFLAISLCAAVGAQSLFIVERGMARNYVFLICALCFMCDIVLMSGVFGVGYF- 62

QY 66 AAPVILDI-MRWGGIAYLLWEAVMAAKDAMTKNKEAPQIIESTPTVDDTPLGGSATV 124  
 Db 63 AKNLYLSLFNLFAGVTFGTGYAFLAK-----T 90  
 QY 125 DTRNRVRVEVSDKQVWVKPMLMAVLTLNPNAYLDAFVFIGGVAQVGDTGRWIFA 184  
 Db 91 LFQTEKKKQVQTPKKLSLKKTLTLFTLGVTLNPNQVYLEWVFLIGASAMSFNLVQKVF 150  
 QY 185 GAFPAASLIWFPL-----VGFGAALSRPLSSPKVRWVWVAVVMTALAIAKL 232  
 Db 151 GTLSAASFWMILLCTMSLYG----SKLNNQKIPMGVNLFTVTAINGTSLVTL 199

RESULT 8  
 AAW20968  
 ID AAW20968 standard; Protein; 215 AA.  
 XX AAW20968;  
 XX  
 DT 21-JUL-1997 (first entry)  
 XX  
 DE H. pylori transmembrane protein, hp2p10625orf14.  
 XX  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO9640893-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 06-JUN-1996; 96WO-US09122.  
 XX  
 PR 01-APR-1996; 96US-0630405.  
 PR 07-JUN-1995; 95US-0487032.  
 XX  
 PA (ASTR ) ASTRA AB.  
 XX  
 PI Berglindh OT, Smith D, Mellgaard BL;  
 XX WPI: 1997-052306/05.  
 DR N-PSDB; AAT68221.  
 DR  
 XX Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter  
 XX  
 PS Claim 73; Page 1359; 1481pp; English.  
 XX  
 CC This sequence represents a H. pylori protein likely to contain five  
 CC membrane spanning regions.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.  
 XX  
 SQ Sequence 215 AA;  
 Query Match 14.3%; Score 170.5; DB 18; Length 215;  
 Best Local Similarity 21.5%; Pred. No. 3.7e-11;

Matches 50; Conservative 48; Mismatches 92; Indels 43; Gaps 5;

QY 6 IFITGLLGASLLSIGPQNVLVIKQIKREGIAVLVCLISDVFLFIAGTILGVDLLSN 65  
 Db 9 VFIEFGGLAISLCAVGAQSLFIVERGMARNYVFLICALCFMCDIVLMSGVGVGAYF- 67

QY 66 AAPVLDI-NRWGGIAYLLWFVAVMAAKDAMTKVAPQIIETEPTVPDDTPLGGSNAVAT 124  
 Db 68 AKNLYLSFLNLFCAVPTGYAFALAK-----T 95

QY 125 DTRNRVRVEVSVDKORVWVVKPMLMAI VLTWLNPNAYLDAFVFLGGVCAQYGDTRMIFAA 184  
 Db 96 LFQTFKKKQVQTPKLSLKTLLFTLGVTLNPOVYLEWVFLIGASMSFNLVQKVFELA 155

QY 185 GAFASLIWFLP-----VGFGAALSRLSPKVKRWINVVAVVWTFALAIKL 232  
 Db 156 GTLSAASFWSLLLLCTMSLRG----SKLANNQKIFMGVNLFTVIMGTLSVTL 204

RESULT 9  
 AAG89911  
 ID AAG89911 standard; Protein; 226 AA.

AC AAG89911;  
 XX  
 DT 26-SRP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 3665.  
 XX  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EF1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX WPI; 2001-376931/40.  
 DR N-PSDB; AAH65130.  
 XX  
 XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 PT  
 XX  
 XX Claim 17; SEQ ID NO: 3665; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
 CC sequences from the Corynebacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of corynebacterium bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
 CC from Corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.

SQ Sequence 226 AA;  
 Query Match 11.1%; Score 132.5; DB 22; Length 226;  
 Best Local Similarity 24.2%; Pred. No. 9.3e-07;  
 Matches 58; Conservative 45; Mismatches 98; Indels 39; Gaps 12;

QY 7 FITGILLGASLLSIGPQNVLVIKQIK--REGIAVLVCLISDVFLFIAGTILGVDLLS 64  
 Db 6 FLALFLVWTAIAASPGDLPQIIRLSAKRRDGVLTAVGIMVGNSTWI-IASLLGLSALI 64

QY 65 NAAPIVLDIMRWGGIAYLLWFVAVMAAKDAMTKVAPQIIETEPTVPDDTPLGGSNAVAT 124  
 Db 65 STYPAILNLQLVGGGILTWGIGAVRSWTVR-----STQQAADSQAVENTLVTA 116

QY 125 DTRNRVRVEVSVDKORVWVVKPMLMAI VLTWLNPNAYLDAFVFLGGVCAQY--GDTG-RW 180  
 Db 117 TAAS-----VGWVPAIRSGIATNLSNPKAVL----PFGSVFAQFVRPDMGICW 160

QY 181 IFAAGAF--AASLIWFLVGFGAALSRLSPSS--PKWRWVWVAVVWTFALAIKMLMG 236  
 Db 161 SIFIGVFLTLTGLLWF--VGF--AVLVRKLAAGLTRNGAII DLTGTGIFIGLGMFIFEG 216

RESULT 10  
 ABP39181  
 ID ABP39181 standard; Protein; 173 AA.

AC ABP39181;  
 XX  
 DT 24-JUL-2002 (first entry)  
 XX  
 DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4026.  
 XX  
 KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;  
 KW antibacterial; gene therapy.  
 XX  
 OS Staphylococcus epidermidis.  
 XX  
 PN US6380370-B1.  
 XX  
 PD 30-APR-2002.  
 XX  
 PF 13-AUG-1998; 98US-0134001.  
 XX  
 PR 14-AUG-1997; 97US-055779P.  
 PR 08-NOV-1997; 97US-064964P.  
 XX  
 XX (GENO-) GENOME THERAPEUTICS CORP.  
 PA  
 XX Doucette-Stamm LA, Bush D;  
 PI  
 XX WPI; 2002-381255/41.  
 DR N-PSDB; ABN91726.  
 XX  
 XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis  
 PT polypeptide, useful for diagnosing and treating bacterial infections -  
 PT  
 XX  
 XX Disclosure; SEQ ID 4026; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading  
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences  
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have  
 CC antibacterial activity and can be used in gene therapy. The sequences  
 CC can also be used in the diagnosis and treatment of bacterial infections,  
 CC particularly S. epidermidis infections. The sequences can be used to  
 CC screen for compounds able to interfere with the S. epidermidis life  
 CC cycle or inhibit S. epidermidis infection.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC USPTO web site.

XX Sequence 173 AA;  
 SQ

CC yield or to provide earlier germination or maturation. The proteins can  
also be used to control insects, to impart resistance to environmental  
stresses, e.g., cold, and to improve nutritional value, e.g., altered oil  
content. The same effects can be produced by producing transgenic plants  
or seeds by incorporation of DNA that encodes A. vitis HR elicitor  
proteins. Use of A. vitis HR elicitor proteins, or nucleic acids encoding  
them, may allow control of previously untreatable diseases; provide  
systemic treatment; and eliminate the need for biological control agents  
or polluting chemicals.

XX CC

SQ Sequence 229 AA;

Query Match 9.6%; Score 114.5; DB 21; Length 229;  
Best Local Similarity 23.3%; Pred. No. 0.00011;  
Matches 59; Conservative 42; Mismatches 79; Indels 73; Gaps 13;

QY 7 FITGILLGASL-----LLSIGPQNVLVIKQGI---KREGLIATLVLCILSDVFLFIAGTIG 59  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 20 FMEQILLIASITWLAVLSPGADFANVRGRNLFYGRKSGLAASMGIA-IACWFHVIYAMFG 78  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 60 VDLLSNAAPIVLIDIMRWGGIAYLLWFAFMAAKOAMTNKVAPQIIBETEPTVPDDTPLG 119  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 79 IAIQHIFPNILDIKFVGAAIV-YAGLATAFSKIEDVEG-----SLVPSDRSMG- 128  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 120 SAVATDTRNRVRVESVDQRVVVKPMLMAIVLTWLMNPAYLDFAVFVGGAQY--GDT 177  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 129 -----RENMGTGLTNGLNPK----TSIFVISLYTFQTIGKDT 160  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 178 G-----RWIFAAGAF--AASLIWFPLVGFGAALSRLLSPKV-----WRWINVVAV 223  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 161 PLUSHQLW---GLFISHLWLNF-----ASVSTLSNPARTVTLRRQRLFNILIG 209  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 224 VMTALAIKMLMG 236  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
Db 210 VLASIGAILFTAG 222  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 12  
AAY99598  
ID AAY99598 standard; Protein; 206 AA.  
AC AAY99598;  
XX XX  
DT DT  
DE DE  
DE DE  
KW KW  
KW KW  
XX XX  
OS OS  
XX XX  
PN PN  
XX XX  
PD PD  
XX XX  
PF PF  
XX XX  
PR PR  
XX XX  
PA PA  
PI PI  
XX XX  
XX XX  
DR DR  
DR DR  
XX XX  
PT PT  
PT PT  
PT PT  
XX XX  
XX XX

08-SEP-2000 (first entry)

E. coli L-threonine resistance protein, RhtC.

L-threonine resistance; L-threonine synthesis; rhtC;  
L-homoserine; L-valine; L-leucine.

Escherichia coli.

EP1013765-A1.

28-JUN-2000.

20-DEC-1999; 99EP-0125406.

23-DEC-1998; 98RU-0123511.

(AJIN ) AJINOMOTO KK.

Livhits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;  
WPI; 2000-414602/36.  
N-PSDB; AAA48443.

Novel Escherichia bacterium having enhanced L-threonine resistance due  
to enhanced RhtC protein activity, used to produce L-threonine,  
L-homoserine, L-valine and L-leucine -

Claim 1; Page 15-16; 24pp; English.



CC The present sequence is the L-threonine resistance protein, RhtB, from  
 CC Escherichia coli. The coding sequence may be used to impart L-threonine  
 CC resistance on E. coli bacteria, which would be useful for producing  
 CC a high yield of L-threonine. L-threonine resistance means that the  
 CC bacteria will be able to grow on a minimal medium containing  
 CC L-threonine at a concentration at which the corresponding wild-type  
 CC strain would not grow. Since the transformed bacteria can grow on the  
 CC minimal medium, it can synthesize L-threonine, which accumulates. The  
 CC accumulated amino acids can then be removed from the culture medium.  
 CC The bacterium of the present invention may also be used to synthesize  
 CC L-homoserine, L-valine and L-leucine at increased levels.

XX  
 SQ Sequence 206 AA;

Query Match 8.8%; Score 105; DB 21; Length 206;

Best Local Similarity 20.4%; Pred. No. 0.0012;

Matches 51; Conservative 36; Mismatches 99; Indels 64; Gaps 9;

QY 3 IMEFTITGLLGLASLLISIGPQNVLVTKQIKR---EGLTAVLLVCLISDVFLFIAGTIG 59

DB 1 MLMLFLTVAMVHVALMSPGDFFFVSQTAVSRKEMMGVLGITCGVMVWAGIA-LLG 59

QY 60 VDLISNAPIVLDIMRW-----GGIAYLLWFAVMAAKDAMTK---VEAPQIIEETEP 109

DB 60 LHL-----LIEKMAWHLIMVGGGLYLCWGYQMLRGALKKEAVSAPQV----- 106

QY 110 TVPDDTPLGGSATDTRNRVRVEVSDKORVWVKPMLMAIVLTWLPNAYDLDAFVIGG 169

DB 107 -----ELAKSGRS-----FLKGLLTLANPK---AIIYFGS 134

QY 170 -----VGAQYGDTRWTFACAFASLIWFLYGFAGAAALSRPLSSPKWVRWINVVAVV 224

DB 135 VFSLFVGDNGVTARWGIFALLIIVETLAWFTVVASLPALPQMRGORYOLAKWIDGFAGAL 194

QY 225 MTALAIKML 234

DB 195 FAGFGIHLII 204

RESULT 13

AAB01787

ID AAB01787 standard; Protein; 212 AA.

AC AAB01787;

DT 03-JAN-2001 (first entry)

XX Escherichia coli Yeas amino acid excretion protein.

DE E. coli; yeas gene; amino acid production; excretion protein gene;

KW amino acid excretion protein.

XX Escherichia coli.

OS

PN EF1016710-A2.

XX

PD 05-JUL-2000.

XX

PF 17-DEC-1999; 99EP-0125263.

XX

PR 30-DEC-1998; 98RU-0124016.

XX

PR 09-MAR-1999; 99RU-0104431.

XX

XX (AJIN ) AJINOMOTO CO INC.

PA

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;

PI Tokhmakova IL;

XX

DR WPI; 2000-414802/36.

XX

DR N4PSDB; AAA52689.

XX

PT Increased production of L-amino acids by an Escherichia bacterium  
 PT comprises increasing the expression amount of an L-amino acid excretion

PT protein -

XX Claim 1; Page 21; 29pp; English.

XX

CC The present sequence is the Yeas amino acid excretion protein from  
 CC Escherichia coli. This protein is involved in the production of amino  
 CC acids, and an increase in its expression leads to an increased  
 CC accumulation of amino acids in the cell. In this case, an increase in  
 CC lysine, alanine, valine, histidine, isoleucine, glutamic acid and proline  
 CC is achieved if multiple copies of its gene are transfected into a  
 CC bacterium. The bacterium used is E. coli.

XX  
 SQ Sequence 212 AA;

Query Match 8.5%; Score 101; DB 21; Length 212;

Best Local Similarity 22.3%; Pred. No. 0.0036;

Matches 52; Conservative 28; Mismatches 69; Indels 84; Gaps 9;

QY 12 LIGASLLLSI-GPQNVLVTKQIK---REGLTAVLLVCLISDVFLFIAGTIGVLLISNAA 67

DB 14 LVGAIFIVLVPENTFLVKNSVSGMGKGYLAACGVFIGDAVLMFLAWA-GVATLIKTT 72

QY 68 PIVLDIMRWGGIAYLLW-----FAVMAAKDAMTKNKEAPQIIEETEPVDDTPLGSSA 121

DB 73 PILFNIVRYLGFAPYLLYLGSKILYATLKGN-----EAKSDEPQYGAI 116

QY 122 VATDTRNRVRVEVSDKORVWVKPMLMAIVLTWLPNNA---YLDAPVFVIGGVGAQYGDGTG 178

DB 117 -----FKRALILSLTNPKAILFYVSPFVQFIDVNAHTGIS 152

QY 179 RMIFRA-----GAPAA-----SLIWFPLVFGA 201

DB 153 PFIILATLELVFCVLSFLIISGAFVQYVIRTKKLVKGVNSLIGLMFVGFPA 205

RESULT 14

AAS58229

ID AAS58229 standard; Protein; 390 AA.

XX

AC AAS58229;

DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #19125.

DE SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

OS

PN WO200181581-A2.

XX

PD 01-NOV-2001.

XX

PF 20-APR-2001; 2001WO-US12865.

XX

PR 21-APR-2000; 2000US-199047P.

XX

PR 02-JUN-2000; 2000US-208841P.

XX

PR 07-JUL-2000; 2000US-216747P.

XX

XX (CORI-) CORIXA CORP.

PA

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX

DR WPI; 2001-616774/71.

XX

DR N-PSDB; AAS59589.

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -

XX PS Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 XX CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 390 AA;

Query Match 8.0%; Score 95; DB 22; Length 390;  
 Best Local Similarity 23.3%; Pred. No. 0.04; 84; Indels 104; Gaps 16;  
 Matches 69; Conservative 39; Mismatches 84; Indels 104; Gaps 16;

QY 1 MVIMEIFITGLGASLLSIGPQNV-----LVKQIKREGLIIV----- 41  
 Db 51 VLLIVFVTVLLFVLGTALSIGVQVPTWVKVVAHVPEGVISVETWPAIDRIVADT 110  
 QY 42 ----LVCLISDVFLTAGLGVLLSN--AAPVLD-----IKRWG----- 77  
 Db 111 RLPRVLLAAVAGMALTTVGTVVQALLRNPLANFTILGVSSGAATGAIANVRFGLLIGTV 170  
 QY 78 -----GIAYL-LWFAVMAAKDANKVAPQIIETEPTVPDTPPLGSAVAT--DTR 127  
 Db 171 SLGLAAGGAFLLLVIMVARQQT--MTAGTLI-----LTGTAVSALLSAV 216  
 QY 128 NRVREVSVDKQVWVKPMLMAIYLTWLPNAYLDAFVFIGVGAQYGDGRW---IFAA 184  
 Db 217 NNFVLTSPD-----PQLASQVLFW-----SLGGFGA-----AKWENLLFPS 253  
 QY 185 GAFA-----ASLWFFPLVG-FGAALSRPLSSPKVWRWVNVVAVMTALAI 230  
 Db 254 GVLAIGIILCLAQASNLNILLAGESAGSLGLNVNRFRTWFA-VAAAIYGVTVAV 308

RESULT 15  
 AAU34590  
 ID AAU34590 standard; Protein; 452 AA.  
 XX AAU34590;  
 XX 14-FEB-2002 (first entry)  
 XX E. coli cellular proliferation protein #171.  
 XX Antisense; prokaryotic cellular proliferation protein;  
 KW antibiotic; antibacterial; drug design.  
 XX Escherichia coli.  
 OS  
 XX WO200170955-A2.  
 XX 27-SEP-2001.  
 XX 21-MAR-2001; 2001WO-US09180.  
 PF 21-MAR-2000; 2000US-191078P.  
 XX

PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR N-PSDB; AAS52449.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 10183; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 452 AA;

Query Match 7.9%; Score 93.5; DB 22; Length 452;  
 Best Local Similarity 19.4%; Pred. No. 0.073; 95; Indels 77; Gaps 14;  
 Matches 52; Conservative 44; Mismatches 44; Indels 77; Gaps 14;

QY 8 ITGLLLG---ASLLSIGPQNVLV-----IKQIKR--EGLIYLVCLISDVFLFIAGT 57  
 Db 150 ISGIIIGLVVAEMFTFIVRRNVWIKLPDSVPASVRSFSAIPGFIILSVMGIIAVALNT 209  
 QY 58 LG-----VDLLSNAPVLDIMRWGGIAY--LWFAVM---AKDANKVAPQIIE 105  
 Db 210 WGTNFHQIIMDTISTPLASIGSVGVWYVIFVPELLWFFGIHGALALTALDNGIMTPWALE 269  
 QY 106 ETEPTVPDDTPLGGSVAT--DTRNRVREVSVDKQ-RVWVKPMLMAIYLTWLPNAYLDA 163  
 Db 270 N-----IATYQQYGSVEAALAAGKTFHIWAKPM-----LDS 300  
 QY 164 FVFPIGGVG-----AQYGDTRWIFAAFAASLIWFLPVLVFGGAALSRPL 208  
 Db 301 FIFLGGSGATGLILAIPTASRRADYRVAKLPSGIFQ---INEPIL-FGLPIIMNV 356  
 QY 209 SSPKVRWVNVVAVVMTALAIKMLMG 236  
 Db 357 MFIFP-----VLVQFILAAILTAAYMG 379

Search completed: April 24, 2003, 18:22:24  
 Job time : 38 secs



GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:23:28 ; Search time 17 Seconds  
(without alignments)  
1112.390 Million cell updates/sec

Title: US-09-105-117K-2

Perfect score: 1191

Sequence: 1 MVIMEIFITGLLGASLLLS.....INVVAVVMTALAIKLMAG 236

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.\*

```
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1191	100.0	236	9	US-09-738-626-6955
2	1178	98.9	233	9	US-09-746-660A-52
3	132.5	11.1	226	9	US-09-738-626-3665
4	93.5	7.9	452	10	US-09-815-242-10183
5	93	7.8	459	9	US-09-738-626-6998
6	91.5	7.7	446	9	US-09-738-626-6356
7	91.5	7.7	446	10	US-09-951-780-2
8	91	7.6	624	9	US-09-738-626-4289
9	89.5	7.5	205	10	US-09-927-395-2
10	89.5	7.5	205	10	US-09-847-392-2
11	89.5	7.5	340	10	US-09-815-242-11980
12	86	7.2	295	10	US-09-815-242-10104
13	86	7.2	400	9	US-09-738-626-4488
14	84.5	7.1	415	9	US-09-976-059-20
15	83	7.0	295	10	US-09-815-242-14000
16	82	6.9	708	9	US-09-738-626-5024
17	81.5	6.8	478	9	US-09-738-626-3783
18	81.5	6.8	516	9	US-09-738-626-5086
19	80.5	6.8	483	9	US-09-738-626-3588

20	78.5	6.6	452	10	US-09-815-242-13983	Sequence 13983, A
21	78	6.5	446	9	US-10-217-096-4	Sequence 4, Appli
22	76.5	6.4	632	10	US-09-919-497-75	Sequence 75, Appl
23	76	6.4	323	9	US-09-738-626-5239	Sequence 5239, Ap
24	75.5	6.3	386	10	US-09-925-301-1277	Sequence 1277, Ap
25	75.5	6.3	332	9	US-09-738-626-4243	Sequence 4243, Ap
26	75.5	6.3	501	10	US-09-934-868-56	Sequence 56, Appl
27	74.5	6.3	361	9	US-10-028-072-252	Sequence 252, App
28	74.5	6.3	361	9	US-10-121-049-252	Sequence 252, App
29	74.5	6.3	361	9	US-10-123-904-252	Sequence 252, App
30	74.5	6.3	361	9	US-10-140-470-252	Sequence 252, App
31	74.5	6.3	361	9	US-10-175-746-252	Sequence 252, App
32	74.5	6.3	361	9	US-10-176-918-252	Sequence 252, App
33	74.5	6.3	361	9	US-10-176-921-252	Sequence 252, App
34	74.5	6.3	361	9	US-10-137-865-252	Sequence 252, App
35	74.5	6.3	361	9	US-10-140-474-252	Sequence 252, App
36	74.5	6.3	361	9	US-10-142-431-252	Sequence 252, App
37	74.5	6.3	361	9	US-10-143-114-252	Sequence 252, App
38	74.5	6.3	361	9	US-10-140-002-252	Sequence 252, App
39	74.5	6.3	361	9	US-10-142-419-252	Sequence 252, App
40	74.5	6.3	361	9	US-10-123-262-252	Sequence 252, App
41	74.5	6.3	361	9	US-10-142-423-252	Sequence 252, App
42	74.5	6.3	361	9	US-10-121-050-252	Sequence 252, App
43	74.5	6.3	361	9	US-10-141-755-252	Sequence 252, App
44	74.5	6.3	361	9	US-10-143-032-252	Sequence 252, App
45	74.5	6.3	361	9	US-10-123-108-252	Sequence 252, App

#### ALIGNMENTS

##### RESULT 1

```
US-09-738-626-6955
; Sequence 6955, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent in ver. 3.0
; SEQ ID NO 6955
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6955
```

Query Match 100.0%; Score 1191; DB 9; Length 236;

Best Local Similarity 100.0%; Pred. No. 3.4e-112;

Matches 236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVIMEIFITGLLGASLLLSIGPQNVLVIKQIKREGLIALLVCLISDVFLFIAGTLGV 60

Db 1 MVIMEIFITGLLGASLLLSIGPQNVLVIKQIKREGLIALLVCLISDVFLFIAGTLGV 60

QY 61 DLLSNAAPIVLDINWGGIAYILLWFAVMAAKDMTKVPAQIIEETPTVDDTFLGGS 120

Db 61 DLSNAPIVLDIMRWGGIAYLWFAWAAKADMTNKVEAPQIIETEPTVPDDTPLGGS 120  
 QY 121 AVATDTRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRW 180  
 Db 121 AVATDTRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRW 180  
 QY 181 IFAAGAPAAALIFPLVGFCAALSRPLSSPKVWRINNVVAVVMTALAKMLMG 236  
 Db 181 IFAAGAPAAALIFPLVGFCAALSRPLSSPKVWRINNVVAVVMTALAKMLMG 236

## RESULT 2

US-09-746-660A-52  
 ; Sequence 52, Application US/09746660A  
 ; Publication No. US20030049804A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pompejus, Markus  
 ; APPLICANT: Kroger, Burkhard  
 ; APPLICANT: Schroder, Hartwig  
 ; APPLICANT: Zelder, Oskar  
 ; APPLICANT: Haberhauer, Gregor  
 ; APPLICANT: Kim, Jun-won  
 ; APPLICANT: Lee, Heung-Schick  
 ; APPLICANT: Hwang, Byung-Joon  
 ; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
 ; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
 ; FILE REFERENCE: BGI-121CPT  
 ; CURRENT APPLICATION NUMBER: US/09/746,660A  
 ; CURRENT FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 09/606740  
 ; PRIOR FILING DATE: 2000-06-23  
 ; PRIOR APPLICATION NUMBER: 09/603124  
 ; PRIOR FILING DATE: 2000-06-23  
 ; PRIOR APPLICATION NUMBER: 60/141031  
 ; PRIOR FILING DATE: 1999-06-25  
 ; PRIOR APPLICATION NUMBER: 60/142101  
 ; PRIOR FILING DATE: 1999-07-02  
 ; PRIOR APPLICATION NUMBER: 60/148613  
 ; PRIOR FILING DATE: 1999-08-12  
 ; PRIOR APPLICATION NUMBER: 60/189790  
 ; PRIOR FILING DATE: 2000-03-09  
 ; PRIOR APPLICATION NUMBER: DE 19931420.9  
 ; PRIOR FILING DATE: 1999-07-08  
 ; NUMBER OF SEQ ID NOS: 125  
 ; SOFTWARE: PatentIn Vers. 2.0  
 ; SEQ ID NO 52  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-746-660A-52

Query Match 98.9%; Score 1178; DB 9; Length 233;  
 Best Local Similarity 100.0%; Pred. Nq. 6.7e-111;  
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 MEIFITGLLGASLLLSIGPQNVLVIKQIKREGLIAVLVCLISDVFLFIAGTLGVDLL 63  
 Db 1 MEIFITGLLGASLLLSIGPQNVLVIKQIKREGLIAVLVCLISDVFLFIAGTLGVDLL 60  
 QY 64 SNAAPVLDIMRWGGIAYLWFAWAAKADMTNKVEAPQIIETEPTVPDDTPLGGSVA 123  
 Db 61 SNAAPVLDIMRWGGIAYLWFAWAAKADMTNKVEAPQIIETEPTVPDDTPLGGSVA 120  
 QY 124 TDTNRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRWIFA 183  
 Db 121 TDTNRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQYGDTRWIFA 180  
 QY 184 AGAPAAALIFPLVGFCAALSRPLSSPKVWRINNVVAVVMTALAKMLMG 236  
 Db 181 AGAPAAALIFPLVGFCAALSRPLSSPKVWRINNVVAVVMTALAKMLMG 233

## RESULT 3

US-09-738-626-3665  
 ; Sequence 3665, Application US/09738626  
 ; Publication No. US20020197605A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: NAKAGAWA, SATOSHI  
 ; APPLICANT: MIZOGUCHI, HIROSHI  
 ; APPLICANT: ANDO, SEIKO  
 ; APPLICANT: HAYASHI, MIKIRO  
 ; APPLICANT: OCHIAI, KEIKO  
 ; APPLICANT: YOKOI, HARUHIKO  
 ; APPLICANT: TATEISHI, NAKO  
 ; APPLICANT: SENOH, AKIHIRO  
 ; APPLICANT: IKEDA, MASATO  
 ; APPLICANT: OZAKI, AKIO  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-125  
 ; CURRENT APPLICATION NUMBER: US/09/738,626  
 ; CURRENT FILING DATE: 2000-12-18  
 ; PRIOR APPLICATION NUMBER: JP 99/377484  
 ; PRIOR FILING DATE: 1999-12-16  
 ; PRIOR APPLICATION NUMBER: JP 00/159162  
 ; PRIOR FILING DATE: 2000-04-07  
 ; PRIOR APPLICATION NUMBER: JP 00/280988  
 ; PRIOR FILING DATE: 2000-08-03  
 ; NUMBER OF SEQ ID NOS: 7059  
 ; SOFTWARE: PatentIn ver. 3.0  
 ; SEQ ID NO 3665  
 ; LENGTH: 226  
 ; TYPE: PRT  
 ; ORGANISM: Corynebacterium glutamicum  
 US-09-738-626-3665

Query Match 11.1%; Score 132.5; DB 9; Length 226;  
 Best Local Similarity 24.2%; Pred. Nq. 1.2e-05;  
 Matches 58; Conservative 45; Mismatches 99; Indels 39; Gaps 12;

QY 7 FITGLLGASLLLSIGPQNVLVIKQIK--REGLIAVLVCLISDVFLFIAGTLGVDLLS 64  
 Db 6 FLALFLVTAATASPGDLPQIIRLSAKNRDRGVLTAVGIMVGNISWI--IASLLGLSALI 64  
 QY 65 NAAPIVLDIMRWGGIAYLWFAWAAKADMTNKVEAPQIIETEPTVPDDTPLGGSVA 124  
 Db 65 STYPALNLLQLVGGYLTWVGIGAVRSWTKR-----STQQAADSQAVENTLVTA 116  
 QY 125 DTRNRVRVESVDKQVRVKPMLMAVLVTLWLPNAYLDAFVFTGGVGAQY--GDTG-RW 180  
 Db 117 TAAS-----VGWPAIRSGIATNLSNPKAVL----FFGSVPAQFVRPDMGIGW 160  
 QY 181 IFAAGAP--AASLIWPLVGFCAALSRPLSS--PKVWRINNVVAVVMTALAKMLMG 236  
 Db 161 SIFIGVFLTLTGLLWF--VGF--AVLVKLAAGLTRNGAIIDLLTGTVIFIGLGMFIFEG 216

## RESULT 4

US-09-815-242-10183  
 ; Sequence 10183, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078

```

: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,849
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: fastseq for Windows Version 4.0
: SEQ ID NO 10183
: LENGTH: 452
: TYPE: PRT
: ORGANISM: Escherichia coli
US-09-815-242-10183

```

```

Query Match      7.9%; Score 93.5; DB 10; Length 452;
Best Local Similarity 19.4%; Pred. No. 0.25;
Matches 52; Conservative 44; Mismatches 95; Indels 77; Gaps 14;
```

Qy	8	ITGLLG---ASLLISIGPQNVLV-----IKOGIKR--EGLIATLVCLISDVFLFIAGT	57
Db	150	ISGIIIGLVAEMFTFVRNRNVIKLPDSVPASVSSEFSALIPGGIIILVMGIINAWALT	209
Qy	58	LG-----VDLLSNDAPIVLDIRWGGIAY--LLWFAM--AAKDANTNKKVEAPQIE	105
Db	210	WGTFHQIIMDTISTPLASLSGVGWYIVFPVLLWFEGIHGALALTALDNGIMTFPALE	269
Qy	106	ETEPTVPDDTLPGGSAVAT-DTRNKRVREVSVDKQ-RVVVKPMLMAIVLTWLNPNAYLDA	163
Db	270	N-----IATYQQYSVEAALAAGKTFHIWAKPM-----LDS	300
Qy	164	FVFETGGVG-----ROYGDGTGRWIFPAAGAFAASLWFWPLVGFGAALSRL	208
Db	301	FIFLGGGATGLILAIFIASRRADYROVALUPSGIFQ---INEPIL-FGLPIIMNPV	356
Qy	209	SSPKVRWRWINVVAVVTALAIKLMLMG	236
Db	357	MFIFE-----VLVOPIIAAITLAAYYM	376

```

RESULT 5
US-09-738-626-6998
Sequence 6998, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIRO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OKAZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: Patent ver. 3.0

```

```

; SEQ ID NO 6998
;
; LENGTH: 459
;
; TYPE: PRT
;
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6998

```

Query Match	7.8%	Score 93;	DB 9;	Length 459;
Best Local Similarity	23.4%	Pred. No. 0.29;		

Qy	6	IFITGILLGASL--LLSIGPQNVIV-----KQIKREGLIAVL--LVCLISDVFLFIAGT	57
Db	106	VFSFIIIGSIVHEMRNIALTVTVLLVPEAREDKANGLVAGVQGVFLVTSV---IAGS	162
Qy	58	LGVDLLSNAAPIVLIDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQII---EETEPTVDD-	114
Db	163	-ATGFLG-----WEITLW-----ICGLSIVALLHLPIRVDEPELITQEDAQPTVSDDS	211
Qy	115	--TPLGGSAVAT---DTRNRVRVEVSVDKVRVWKPMPLMAIVLTWLNPNAYLDFAVFVIG	169
Db	212	VPTPTSLDAIVSKGIDLKSMKIILSV-----PGLLALVL-----FASFNNLIGG	256
Qy	170	VGAQYGDGTGRWIFAAGAFASLWFLPVGPF-----GRAALSRP---LSSPKVWEWINV	219
Db	257	VYSALMDP-----VGLLELPSPQL-WGLLGLTSLGPIVGGAVISKTGLGKNPVRTLLLVNV	311
Qy	220	VVAVVMTALAIK-----LMLMG	236
Db	312	GVAFVGMFLFAIREWWMWLYIIG	332

```

RESULT-6
US-09-738-626-6356
; Sequence 6356, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6356
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6356

```

	Query Match	7.7%	Score 91.5;	DB 9;	Length 446;
	Best Local Similarity	24.8%;	Pred. No. 0.4;		
	Matches 51;	Conservative 29;	Mismatches 71;	Indels 55;	Gaps 8;
Qy	28	VKQIKREGIIAVLLVCLISDVFLFIAGTGTGVDLLSNAAPIVLIDIMRWGGIAYILLMFAY	87		
Db	234	VQLGLG---ILAFYVTCVI-----FIFGLGVALVKFTGTGVIKFLKYLAKPELLIFAT	285		
Qy	88	MAAKDAMTNKVEAPQIIIEETPTVPDDTPLGGSAVATDTRNRVEVEVSVDKORVWVKPML	147		

```
Db 286 SSSSALPNLMRKMEHGVAKPTGIVVPTGYSFNLGDT-----324
QY 148 MAIVLTWLNPNAYLDAPVFVIG-----GVGAQYGDTRWIFAA-GAFAASLIWFFPLVGF 199
Db 325 -AIVLTW-----ASIFADANMPMSLGEQVGLLVFMIIASKGAAGVS-----GA 368
QY 200 GAALSRLSS--PKWVRWINVVAV 223
Db 369 GIATLAAGLSSHRPELLHGVDDVIGI 394
```

## RESULT 7

```
US-09-951-780-2
; Sequence 2, Application US/09951780
; Patent No. US20020106759A1
; GENERAL INFORMATION:
; APPLICANT: FARWICK, MIKE
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES CODING FOR THE dcta GENE
; FILE REFERENCE: 032301 WD 199
; CURRENT APPLICATION NUMBER: US/09/951,780
; CURRENT FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 446
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-951-780-2
```

```
Query Match 7.7%; Score 91.5; DB 10; Length 446;
Best Local Similarity 24.8%; Pred. No. 0.4;
Matches 51; Conservative 29; Mismatches 71; Indels 55; Gaps 8;
```

```
QY 28 VIKQIKREGIIAVLLVCLISDVFPLFIAGTIGVLDLSNAAPVLDIMRWGGIAYLLWFAV 87
Db 234 VVQLGIL---ILAFVTCVI-----FIFGVLGAVLKVFTGWNIFKVLKYLAKFLLIFAT 285
QY 88 MAKDAMTNKVEAPQIIEETPTVDDTPLGGSAAVATDTRNRVRVSVSKQVRVVKPML 147
Db 286 SSSSALPNLMRKMEHGVAKPTGIVVPTGYSFNLGDT-----324
QY 148 MAIVLTWLNPNAYLDAPVFVIG-----GVGAQYGDTRWIFAA-GAFAASLIWFFPLVGF 199
Db 325 -AIVLTW-----ASIFADANMPMSLGEQVGLLVFMIIASKGAAGVS-----GA 368
QY 200 GAALSRLSS--PKWVRWINVVAV 223
Db 369 GIATLAAGLSSHRPELLHGVDDVIGI 394
```

## RESULT 8

```
US-09-738-626-4289
; Sequence 4289, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
```

```
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4289
; LENGTH: 624
; TYPE: PR1
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4289
```

```
Query Match 7.6%; Score 91; DB 9; Length 624;
Best Local Similarity 20.7%; Pred. No. 0.7;
Matches 54; Conservative 40; Mismatches 71; Indels 96; Gaps 14;
QY 12 LIGASILL-----LSI--GPQNVLVIKQIKREGIIAVLLVCLISDVFLFIAGTIGV 60
Db 112 MLGAALFYGDVVITPAISVLSATEGLTVISPSFER-FILPVSILAVLIA---IPAIOPLGT 167
QY 61 DLLSNA-APVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIEETPTVDDTPLGG 119
Db 168 EKVGKAFGPIM-----LLNEVTLAGL-GIPQIIICHPILQSLSPH-----206
QY 120 SAVATDTRNRVRVSVSKQVRVVKPMLMAIVLTWLNPNAYLDAPVFVIG-----GVGA 172
Db 207 -----WALRLIVA-----EP---FOAFVILGAVVLTVTGAEA 235
QY 173 QYGDTRWIFAAFAASLIWFFLV-----GFGAALSRLPS-----SPKVMR 215
Db 236 LVADMGHF---GARPIRVAMFCVWMPALLITVLGQALVINQPEAVRNPMPFLAPEGLR 291
QY 216 WINVVAVVMTALAIKMLMG 236
Db 292 IPLVILATITATVIASQAVISG 312
```

## RESULT 9

```
US-09-927-395-2
; Sequence 2, Application US/09927395
; Patent No. US20020058314A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENIAMOVICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKMAKOVA, IRINA LVOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERES ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/927,395
; CURRENT FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PR1
; ORGANISM: Escherichia coli
US-09-927-395-2
```

```
Query Match 7.5%; Score 89.5; DB 10; Length 205;
Best Local Similarity 20.6%; Pred. No. 0.23;
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;
QY 4 MEIFITGILLGASLLSIGPQNVLVIKQIKREGIIAVLLVCLISDVFF--LFIAGTIGVD 61
Db 3 LEWWFAYLTSIITLSPGSGAINTMTTSL-NHGYPAGGVYCWASDRTGDSYCAGRWGVG 61
QY 62 LLSNAAPIVLDIMRWGGIAYLLWFAVMAAKDA-----MTN 96
```

```
Db 62 TLFERSVIAEVLKWAAGAYLWLGICQWRAAGAILDKSLASTQSRRLHFORAVFVNLTN 121
QY 97 KVEA-----PQIETEPTVPDDTPLGGSAAVATDTRNRVRVEVSDKQVRVWK 144
Db 122 PKSIVFLAALFPQFIMPQPOLMQYIVLGVTIIVDIIIVMIGVATLAQRIALWIK 176

RESULT 10
US-09-847-392-2
; Sequence 2, Application US/09847392
; Patent No. US20020102670A1
; GENERAL INFORMATION:
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA
; APPLICANT: ALCOSHIN, VLADIMIR VENVAMIONICH
; APPLICANT: BELAREOVA, ALL VALENTINOVNA
; APPLICANT: TOKHAKOVA, IRINA LYOVNA
; TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM
; TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD
; FILE REFERENCE: 0010-1039-0
; CURRENT APPLICATION NUMBER: US/09/847,392
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 09/396,357
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: RU98118425
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-847-392-2

Query Match 7.5%; Score 89.5; DB 10; Length 205;
Best Local Similarity 20.6%; Pred. No. 0.23;
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;

QY 4 MEIFITGLGASLLSIGPQNVLVIKQIKREGLIAVLVCLISDVFLFIAGTLGVD 61
Db 3 LEWFAVLLTSIILTUSPGSGAINTMTTSL-NHGYAGGYCWASDRGTSYCGMRGVG 61
QY 62 LLSNAAPVLDIMRWGGIAYLLWFAVMAAKDA-----MTN 96
Db 62 TLFERSVIAEVLKWAAGAYLWLGICQWRAAGAILDKSLASTQSRRLHFORAVFVNLTN 121
QY 97 KVEA-----PQIETEPTVPDDTPLGGSAAVATDTRNRVRVEVSDKQVRVWK 144
Db 122 PKSIVFLAALFPQFIMPQPOLMQYIVLGVTIIVDIIIVMIGVATLAQRIALWIK 176

RESULT 11
US-09-815-242-11980
; Sequence 11980, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```

```
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; -SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11980
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11980

Query Match 7.5%; Score 89.5; DB 10; Length 340;
Best Local Similarity 24.4%; Pred. No. 0.44;
Matches 61; Conservative 29; Mismatches 89; Indels 71; Gaps 12;

QY 1 MVIMEIFITGLLIGASLLSIGPQNVLVIKQIKREGLIAVLVCLISDVFLFIAGTLG 59
Db 113 VALFDLASMGQYLGCA-FLAGLAGIAVFLGQARETGTNPNVLVLGAGLSVNLASLTG 171
QY 60 VDLSNAAPVLDIMR-W-----GGIAYLLW-----FAVMAAKDAMTNKVEAPQIIEE 106
Db 172 IIVL-NAPPEVDFRFRHWAAGSLSGSFALLGWPGLAIGAGLAAAFALAAARNALAGQE 230
QY 107 TEPTVPDDTPLGGSAAVATDTRNRVRVEVSDKQVRVWKPMIAVLVTLNPNAYLDAFVP 166
Db 231 I-----GQALGVDLR-----LTWL--LACLAVMLL 253
QY 167 IGVGGAQYGDTRWTFAGAAFAASLIWFFLVGFGAAALSRPLSSPKVWRWVWVAVVMT 226
Db 254 AG-----AATAGAPIAF-----VGLVAPHLARLLAGPD-QRWILPFSALIAA 295
QY 227 ALAKMLMG 236
Db 296 GLLGADILG 305

RESULT 12
US-09-815-242-10104
; Sequence 10104, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```



;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: Fast-Seq for Windows Version 4.0  
;; SEQ ID NO 10104  
;; LENGTH: 295  
;; TYPE: PRT  
;; ORGANISM: Escherichia coli  
US-09-815-242-10104

Query Match 7.2%; Score 86; DB 10; Length 295;  
Best Local Similarity 21.4%; Pred. No. 0.82; Indels 80; Gaps 13;  
Matches 58; Conservative 39; Mismatches 88

QY 1 MVIMEIFITGLLGASLLSITGPONVLVIKQIK--REGLIAVLLVCLISDVFLFIAGTL 58  
DB 14 IVILLVAMASIOGGASLAKSLFP---LVGAPGVTLRLALGTLILIAFFKWLRFKQEQ 70  
QY 59 GVDLLSNAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIBETPTVPDDTPLG 118  
DB 71 RLPLLFYGVSL-----GGNYLFYLSIQ-----TVP-----L 97  
QY 119 GSAVATDTRNRVRVEVSDKQR---VWVKPMLMAIVLTW-----L 155  
DB 98 GIAVALEFTGTLAVALFSRRPVDVFWV---VLAVALGLWFLPLGQDVSHVDLTGCALAL 154  
QY 156 NPNAYLDAPFVIG-GVGAQYDGTGRWIFAAGAPAAASLIWFLVGFVGAALSRPLSSPKW 214  
DB 155 GAGCAWAIYLSQGRAGHGA---TVAIGSLIALIFVPI---GALQAGEAL-----W 203  
QY 215 RW-----INVVVAVMTALAKMLM 235  
DB 204 HWSVIPGLAVAILSTALPKSLEMI 228

RESULT 13  
US-09-738-626-4488  
;; Sequence 4488, Application US/09738626  
;; Publication No. US20020197605A1  
;; GENERAL INFORMATION:  
;; APPLICANT: NAKAGAWA, SATOSHI  
;; APPLICANT: MIZOGUCHI, HIROSHI  
;; APPLICANT: ANDO, SEIKO  
;; APPLICANT: HAYASHI, MIKIRO  
;; APPLICANT: OCHIAI, KEIKO  
;; APPLICANT: YOKOI, HARUHIKO  
;; APPLICANT: TATEISHI, NAKO  
;; APPLICANT: SENOH, AKIHIRO  
;; APPLICANT: IKEDA, MASATO  
;; APPLICANT: OZAKI, AKIO  
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
;; FILE REFERENCE: 249-125  
;; CURRENT APPLICATION NUMBER: US/09/738,626  
;; CURRENT FILING DATE: 2000-12-18  
;; PRIOR APPLICATION NUMBER: JP 99/377484  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: JP 00/159162  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: JP 00/280988  
;; PRIOR FILING DATE: 2000-08-03  
;; NUMBER OF SEQ ID NOS: 7059  
;; SOFTWARE: PatentIn ver. 3.0  
;; SEQ ID NO 4488  
;; LENGTH: 400  
;; TYPE: PRT  
;; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-4488

Query Match 7.2%; Score 86; DB 9; Length 400;  
Best Local Similarity 21.4%; Pred. No. 1.2;  
Matches 54; Conservative 27; Mismatches 73; Indels 98; Gaps 15;

QY 6 IFITGLLGASLLSITGPONVLVIKQIKREGLIAVLLVCLISDVFLFIAGTLG--VDLL 63  
DB 103 IFLTGF-----VVMGIADVAIL-LPAWIKLHGKSTVALMTVYTVLVGVSTLGLPLSTLL 156  
QY 64 ---SN-----RAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIBETEPTV 111  
DB 157 FSGSNARWAIPTWLPVQLVAIN---LPMWW-----DKYDF 192  
QY 112 PDPTPLGGSVATDTRNRVRVEVSDKQRVWVKPMLMAIVL-----TWLNFN 158  
DB 193 PAETVKSST-----GKIWTSPTAFFIMLFGLQSMNAYIQMGWL-PK 234  
QY 159 AYLDAPVFITGGVGAQYDGTGRWIFA---AGAPA-----ASLIWFLVGFQAAL 204  
DB 235 IFIDA-----GVSPAHSIGLSIVGIMVGGTTPVAIARTDKNLVWFPVY-FGASMF 288  
QY 205 SRPLSSPKVWRW 216  
DB 289 ---LGVGTWLV 297

## RESULT 14

US-09-976-059-20  
;; Sequence 20, Application US/09976059  
;; Patent No. US20020164747A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Farnet, Chris  
;; APPLICANT: Zazopoulos, Emmanuel  
;; APPLICANT: Staffa, Alfredo  
;; TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin  
;; FILE REFERENCE: 3019-PCT  
;; CURRENT APPLICATION NUMBER: US/09/976,059  
;; CURRENT FILING DATE: 2001-10-15  
;; NUMBER OF SEQ ID NOS: 34  
;; SOFTWARE: PatentIn version 3.0  
;; SEQ ID NO 20  
;; LENGTH: 415  
;; TYPE: PRT  
;; ORGANISM: Actinoplanes sp.  
;; NAME/KEY: misc.feature  
;; LOCATION: (1)..(1)  
;; OTHER INFORMATION: V represents a non-standard initiator codon. It is expected that  
;; OTHER INFORMATION: the biosynthesized protein will have a formylmethionine residue  
;; OTHER INFORMATION: at this position  
US-09-976-059-20

Query Match 7.1%; Score 84.5; DB 9; Length 415;  
Best Local Similarity 20.6%; Pred. No. 1.8;  
Matches 50; Conservative 37; Mismatches 71; Indels 85; Gaps 10;

QY 1 MVIMEIFITGLL-----GASLLSITGPONVLVIKQIKREGLIAVLLVCLISDVFLF 53  
DB 116 LAVLTWATGAVLRTGRRPGDALLVAAP-----VLI 148  
QY 54 IAGTLGVLLSNAAPITVLDIMRWGG-----IAYLLWFAVMAAKDAMTNKVEAPQIIBETEPT 109  
DB 149 LAGTVNMDLLAVAA-AVLAIALAWERDRPLLAGVLIGLGTAAK-----LFPLVL----- 195  
QY 110 TVPDDTPLGGSVATDTRNRVR--VEVSDKQRVWVKPMLMAIVLTWLNPNAYLDAPVFI 167  
DB 196 -----LGPVLLLCILQRMRFRFAVAAAGAAW---LLVNLVVVALQPDGWMFEFRFN 245  
QY 168 GGVAQYDGTGRWIFAAGAPAAASLIWFLVGFQAALSRPLSSPKVWRWVNVVAVMTA 227  
DB 246 AGCAEFGS-----LWFLDGLG-----LWPAVNAVALATFGVLLAG 283  
QY 228 LAI 230  
DB 284 IAV 286





GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 24, 2003, 18:21:48 ; Search time 15 Seconds  
(without alignments)  
462.921 Million cell updates/sec

Title: US-09-105-117k-2  
Perfect score: 1191  
Sequence: 1 MVIMRIFITGLLGASLLLS.....INNVAVVMTALAIAKMLMG 236

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2.6/prodata/1/iaa/5A COMB.pep.\*  
2: /cgn2.6/prodata/1/iaa/5B COMB.pep.\*  
3: /cgn2.6/prodata/1/iaa/6A COMB.pep.\*  
4: /cgn2.6/prodata/1/iaa/6B COMB.pep.\*  
5: /cgn2.6/prodata/1/iaa/PCUS COMB.pep.\*  
6: /cgn2.6/prodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	119.5	10.0	173	4	US-09-134-001C-4026
2	89.5	7.5	205	4	US-09-396-357-2
3	85.5	7.2	224	4	US-09-134-001C-4608
4	82.5	6.9	3491	2	US-07-642-734C-2
5	82.5	6.9	3491	3	US-08-439-009A-2
6	80	6.7	377	4	US-09-105-390-64
7	79	6.6	478	4	US-09-134-001C-4637
8	79	6.6	530	4	US-08-793-044-3
9	79	6.6	4545	2	US-08-804-227C-14
10	79	6.6	4550	2	US-08-804-227C-8
11	79	6.6	4550	2	US-08-804-198-2
12	78.5	6.6	831	2	US-08-677-734A-11
13	78.5	6.6	831	4	US-09-097-053-11
14	76.5	6.4	834	2	US-08-677-734A-9
15	76.5	6.4	834	2	US-08-677-734A-10
16	76.5	6.4	834	4	US-09-097-053-9
17	76.5	6.4	834	4	US-09-097-053-10
18	74.5	6.3	5215	4	US-09-105-537-2
19	73.5	6.2	484	4	US-09-134-001C-5063
20	73	6.1	303	2	US-08-846-762-20
21	73	6.1	303	2	US-08-846-762-91
22	73	6.1	502	4	US-09-134-001C-5674
23	72.5	6.1	488	2	US-08-928-692-10
24	72.5	6.1	488	4	US-09-339-972-10
25	72	6.0	192	1	US-08-112-208C-9
26	72	6.0	192	1	US-08-248-819A-9
27	72	6.0	192	1	US-08-607-269-25

28	72	6.0	192	1	US-08-471-058-13	Sequence 13, Appl
29	72	6.0	192	2	US-08-337-646A-9	Sequence 9, Appl
30	72	6.0	192	2	US-08-856-531-9	Sequence 9, Appl
31	72	6.0	192	2	US-08-856-034-9	Sequence 9, Appl
32	72	6.0	192	3	US-08-471-057-13	Sequence 13, Appl
33	72	6.0	192	4	US-09-127-048-7	Sequence 7, Appl
34	72	6.0	192	4	US-08-927-326-9	Sequence 9, Appl
35	72	6.0	192	5	PCT-US95-04600-25	Sequence 25, Appl
36	72	6.0	245	4	US-08-469-260A-40	Sequence 40, Appl
37	72	6.0	450	3	US-09-045-631-2	Sequence 2, Appl
38	72	6.0	450	4	US-09-158-843A-2	Sequence 2, Appl
39	72	6.0	1447	4	US-09-041-886-25	Sequence 25, Appl
40	72	6.0	1447	5	PCT-US94-05277-2	Sequence 2, Appl
41	72	6.0	1528	1	US-08-463-092B-6	Sequence 6, Appl
42	72	6.0	1528	2	US-08-463-109A-6	Sequence 6, Appl
43	72	6.0	1528	2	US-08-460-907B-6	Sequence 6, Appl
44	72	6.0	1528	3	US-08-463-179A-6	Sequence 6, Appl
45	72	6.0	1528	3	US-08-461-384B-6	Sequence 6, Appl

ALIGNMENTS

RESULT 1  
US-09-134-001C-4026  
; Sequence 4026, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4026.  
; LENGTH: 173  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4026

Query Match 10.0%; Score 119.5; DB 4; Length 173;  
Best Local Similarity 23.0%; Pred. No. 3.5e-06;  
Matches 46; Conservative 32; Mismatches 83; Indels 39; Gaps 4;  
QY 39 IAVLLVCLISDVFLFTAGTGLGVLDLSNRAPIVLDIMRWGGIAVLLWFAVWAAKDMTKV 98  
Db- 8 LPVITAGUCDTFLIVAILGVSLIISMTLQLFYIIGFLFLMTM-----AWSLWT 60  
QY 99 EAPQIIIEETPTVPDDTPLGGSAAVATDTNRVRVRSVVDKQVWVKPMLMAIVLTWLNPN 158  
Db 61 EXPNSIEETP-----MSAKQ-----ILFALSLSLNLPH 90  
QY 159 AVLDAPVFTGGVGAQYGDTRWIFAGAPAAASLIWPLPVGGAALSRPLSPKVRWTN 218  
Db 91 AIMDTVGVTGSASVVDGVKVFSLATISVSWINFWFLAILGRITGKIDKSGKVIILN 150  
QY 219 VV--VAVVMTALAIAKMLMG 236  
Db 151 KVSSVIVIIIGLIILKNIVG 170  
RESULT 2  
US-09-396-357-2  
; Sequence 2, Application US/09396357  
; Patent No. 6303348  
; GENERAL INFORMATION:  
; APPLICANT: LIVSHITS, VITALY ARKADIEVICH  
; APPLICANT: ZAKATAEVA, NATALIYA PAVLOVNA

APPLICANT: ALCOSHIN, VLADIMIR VENTAMIOVICH  
APPLICANT: BELAREOVA, ALL VALENTINOVA  
APPLICANT: TOKHAKOVA, IRINA LVOVNA  
TITLE OF INVENTION: DNA CODING FOR PROTEIN WHICH CONFERS ON BACTERIUM  
TITLE OF INVENTION: ESCHERICHIA COLI RESISTANCE TO L-HOMOSERINE AND METHOD  
TITLE OF INVENTION: FOR PRODUCING L-AMINO ACIDS  
FILE REFERENCE: 0010-1039-0  
CURRENT APPLICATION NUMBER: US/09/396,357  
CURRENT FILING DATE: 1999-09-15  
EARLIER APPLICATION NUMBER: RU98118425  
EARLIER FILING DATE: 1998-10-13  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: Patent in Ver. 2.1  
SEQ ID NO 2  
LENGTH: 205  
TYPE: PRP  
ORGANISM: Escherichia coli  
US-09-396-357-2

Query Match 7.5%; Score 89.5; DB 4; Length 205;  
Best Local Similarity 20.6%; Pred. No. 0.015;  
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;

QY 4 MEIETITGLLGLSLLSGPQNVLVKQIKREGLIAVLVCLISDVF--LFIAGTLGVD 61  
DB 3 LEWFAVLLTSIILTLSPGSAINTMTSL-NHGYPAGGVTCWASDRGTGDSYCAWGRGVG 61

QY 62 LLSNAAPVILDMRWGGIAYLLWFAVMAAKDA-----MTN 96  
DB 62 TLFERSVIAPEVLKXWAGAVLWLGIGQWRAAGAILDKSLASTQSRHLPQRAVFNLTN 121

QY 97 KVEA-----PQIETEPTVPDDTLPGGSATDTRNRVRVSVDKORVWK 144  
DB 122 PKSIVFLAALFPQFIMPQPOLQMYVLGVTVTVVDDIIVMIGYATLAQRIALMIK 176

RESULT 3  
US-09-134-001C-4608  
Sequence 4608, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4608  
LENGTH: 224  
TYPE: PRP  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4608

Query Match 7.2%; Score 85.5; DB 4; Length 224;  
Best Local Similarity 19.5%; Pred. No. 0.049;  
Matches 48; Conservative 42; Mismatches 103; Indels 53; Gaps 8;

QY 3 IMEITITGLLGLSLLSGPQNVLVKQI---KREGLIAVLV-----CLISDVFIFIA 55  
DB 14 LMDGLITFIITLIIIVPGDFIIVMKNTINSKMGFMFAAFGITTHILYSSLAIF-- 71

QY 56 GTLGVLLSNAAPVILDMRWGGIAYLLWFAVMAAKDAMTKVEAPQIETEPTVPDDT 115  
DB 72 ---GIYILSLHFVFTIKILGACIYILGKISLSAHSVDFSKQAL----- 117

QY 116 PLGGSATDTRNRVRVSVDKORVWKPMMLMAIVLTWLNPNAYLDAFVFIGGGAQYG 175  
DB 118 -----ADVRN-----VITSFQGFSLSLNPKALL-----FYVSIRPQPL 154

QY 176 DTGR-----WIFAAGAFAAASLIWPLVGFAGAAALSRFLSPKVRWVWVVMVMTAL 228  
DB 155 SNGNIHMKSEVALFAPSVVWVICLWELFCVFFIQYIKLFLSPRPKAIKFDYIVGVFLIGL 214

QY 229 AIKML 234  
DB 215 SINLL 220

RESULT 4  
US-07-642-734C-2  
Sequence 2, Application US/07642734C  
Patent No. 5824513  
GENERAL INFORMATION:  
APPLICANT: Katz, L  
APPLICANT: Donadio, S  
APPLICANT: Mcalpine, J B  
TITLE OF INVENTION: Recombinant DNA Method for Producing  
TITLE OF INVENTION: Erythromycin Analogs  
NUMBER OF SEQUENCES: 27  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Edward H. Gorman  
STREET: Abbott Laboratories D377/AP6D-2 One Abbott  
STREET: Park Rd  
CITY: Abbott Park  
STATE: IL  
COUNTRY: US  
ZIP: 60064-3500  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/642,734C  
FILING DATE: 17-JAN-91  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Danckers, Andreas M  
REGISTRATION NUMBER: 32652  
REFERENCE/DOCKET NUMBER: 4952.US.01  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 708-937-9396  
TELEFAX: 708-938-2623  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3491 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-642-734C-2

Query Match 6.9%; Score 82.5; DB 2; Length 3491;  
Best Local Similarity 28.0%; Pred. No. 5.8;  
Matches 52; Conservative 18; Mismatches 53; Indels 63; Gaps 12;

QY 66 AAPIVLDI---MRWGGIAYLLW-----FAVMAAKDAMTKVE-----APQIETEPT 110  
DB 53 SAPVVFVFGQAGWAGMAGELLGSRVFA--AAMDACARAFEPVTDWTLAQVLDSPQOS 110

QY 111 VPDDTLPGGSATDTRNRVRVSVDKORVWKPMMLMAI-----VLTWLNPNAYLDAFV- 165  
DB 111 -----RRVEV-----VQPALFAVQTSLAALWRSFGVTPDAVVG 143

QY 166 -FTGGVGAQY--GDTGRWTFAGAFASLIW-----FPLVGF-----AAALSRPLSSPKVR 215  
DB 144 HSTGEIAAHVCGAAG-----AADAARAALWREMIPLVNGMDMAVALSADEIEPRIAR 199

QY 216 WINVV 221  
DB 200 WDDVV 205



; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4637  
; LENGTH: 478  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4637

Query Match 6.6%; Score 79; DB 4; Length 478;  
Best Local Similarity 19.7%; Pred. No. 0.84;  
Matches 58; Conservative 45; Mismatches 80; Indels 112; Gaps 13;

QY 29 IKQIKREGL---IAVLVCLISDVFIAGTL-GVDLL-----SNAA-----PIVL 71  
DB 102 ISRGVTKLGRRIALQFVKLFGKTLGLAYSILVGVDLILAPATPSNTARAGIMPIIK 161  
QY 72 DIMRWGG-----IA 80  
DB 162 SLSEFGSSPRDGERKMGAFILFTEQGNLITSAMFLTAMAGNPITAOQLAHTAHVQIT 221  
QY 81 YLLWFVMAAKADAMTKVEAPQIETEPTVPDDTP-----LGSSAVATDTRN 128  
DB 222 WMNWF-VAALIPGLISLIVVFFIYKLYPPTVKETPNNAKKWATEQLEEMGHMSIA-----E 276  
QY 129 RVVEVSVDKQRVWVKPMLMAVLVTLNPNAYLDAFV-----FIGGVGAQ---YGDTRW 180  
DB 277 KLMGVFIIALWV-----LGSFINDVATLTAFIALALLLTGLAWSDIILNETGAW 329  
QY 181 IFAGAPASLIWFLVFGGAALSRPLSPKVMRWINVVVMTALAIKMLM 235  
DB 330 -----NTLVWFSVLVMAEQNLKGFIP-----WLSKLIAGLNGFSWPVIV 372

## RESULT 8

US-08-793-044-3  
Sequence 3, Application US/08793044  
Patent No. 6235497

## GENERAL INFORMATION:

APPLICANT: Bejanin, Stephane  
APPLICANT: Berrard, Sylvie  
APPLICANT: Cervini, Riccardo  
APPLICANT: Mallet, Jacques  
TITLE OF INVENTION: NOVEL VESICULAR ACETYLCHOLINE CARRIER  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road, Mailstop 3c43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426

COMPUTER READABLE FORM: disk  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/793,044  
FILING DATE:  
CLASSIFICATION: 536

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR95/01073  
FILING DATE: 10-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/10044  
FILING DATE: 16-AUG-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Savitzky Esq., Martin F.  
REGISTRATION NUMBER: 29,699  
REFERENCE/DOCKET NUMBER: ST94066-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3816  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:  
; LENGTH: 530 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-793-044-3

Query Match 6.6%; Score 79; DB 4; Length 530;  
Best Local Similarity 23.5%; Pred. No. 0.98;  
Matches 52; Conservative 31; Mismatches 74; Indels 64; Gaps 12;

QY 9 TGLLGASLLLSIGPQNVLVIKQIKREGLIAVLVC--LISDVFLFIAGTLGVLLSNA 66  
DB 7 TQOARAAATKLS---EAVGAALQEFQRQRRLVIVCVALLDNMLYV-----I 53  
QY 67 APIVLDI---MRWGGIAYLLWPAVMAAKDAMTKVEAPQIIEET-EPTVPDDTPIGGSAV 122  
DB 54 VPIVPDYIAHMEGGS-----EGPTLVSEVWEPTLPPTLANASAY 93  
QY 123 ATDT-----RNRVREVSVDKQRVWVKPML--MAIVLTWLNPN--NAVLDAPV-- 166  
DB 94 LANTSASPTAAGSARSILRPYPTSEDEVKIGVLFASKAILQLLVNPLSGPFIDRMSYDV 153  
QY 167 --IGGVGAQYGDTRWIPA---AGAPFAASLIWFLVFGGAA 202  
DB 154 PLLIGLVWFASTVWFAPFAEDYATLFAAR-----SLQGLGSA 190

## RESULT 9

US-08-804-227C-14  
Sequence 14, Application US/08804227C  
Patent No. 5876991

## GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.  
APPLICANT: Kuhstoss, Stuart A.  
APPLICANT: Rostock, Paul R., Jr.  
APPLICANT: Sutton, Kimberly L.  
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA  
ZIP: 46285

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: ASCI(DOS) Text only  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-8231  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4545 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-804-227C-14

Query Match 6.6%; Score 79; DB 2; Length 4545;  
Best Local Similarity 21.5%; Pred. No. 22;  
Matches 58; Conservative 35; Mismatches 93; Indels 84; Gaps 11;

	Oy	3	IMEIITGTLGLLGSALLSGPQNVLVIKQGIKREGILAVLLVCLISDVFLFIAGTGLGVDL	62
	Ddb	2163	VLRTPVGAALAGENQFAVRPSGV-----HVVRRVPAPVPVPASARTVTTPATAVGEDA	2217
	Oy	63	LSNAAPIVLIDNRW-----GGIAYLLWFPAVMAAKDAMTKVE-----AP	101
	Ddb	2218	RNDTSDDVVPPDDRWSGSLVTIGTGTA--GAQVARRLARSGAARLLLVRGRGAAGPGV	2275

QY	3	IMEIFITGLLGASLLISIGPQNVLVIKQGIKREGLIANVLVCLISDVFLFIAGTILGVDL	62
Db	2163	VLRTFVGALLAGGENQFAVRPSGV-----HVRVVPAPVPVPASARTVTTAPATAVGEDA	2217
QY	63	LSNAAPIVLIDIRW-----GGIYLLFWFAVMAKDAMTNKVE-----AP	101
Db	2218	RNDTSDVVVPDDRSSSGTVLTIGTGAL--GAQVARELARSGAARLLVGRRGAAQPGVG	2275
QY	102	QIIIES-----TEPTV-----PDDTF-----LGSSAVATD-----TRN	128
Db	2276	ELVVELTALGSEVAECVADRDLAALLAGLFEERPLVAVLHAAGVLDGVDLSLTSD	2335
QY	129	RVRVEVSDKQVRVMKPLMAIVLTINPNAYLIDAFVFIGVGVAQYQDGTGRWTFAA-----	184



Db 2336 RVDA-VLRDK-----VTARHDELTAIDLPLDAFLVLFSSIVGWGNGQAVAAANAA 2387  
 QY 185 -----GAPASLIWFPPLVFGAAA 203  
 Db 2388 LDALQRRRARGARAASIAWGPWAGWAS 2417

RESULT 12  
 US-08-677-734A-11  
 ; Sequence 11, Application US/08677734A  
 ; Patent No. 5871919  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brant, Steven R.  
 ; APPLICANT: Yun, Chris C.H.  
 ; APPLICANT: Donowitz, Mark  
 ; APPLICANT: Tse, Chung-Ming  
 ; TITLE OF INVENTION: Cloning, Tissue Distribution, and  
 ; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,  
 ; TITLE OF INVENTION: NHE3.  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESSEE: Dunner  
 ; STREET: 1300 I Street, N.W., Suite 700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/677,734A  
 ; FILING DATE: 10-JUL-1996  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fordis, Jean B.  
 ; REGISTRATION NUMBER: 32,984  
 ; REFERENCE/DOCKET NUMBER: 05387.0043-00000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 408-4000  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 831 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-08-677-734A-11

Query Match 6.6%; Score 78.5; DB 2; Length 831;  
 Best Local Similarity 20.4%; Pred. No. 2.1;  
 Matches 53; Conservative 35; Mismatches 93; Indels 79; Gaps 11;  
 QY 6 IFITGL-----LLGASLLLSIGPQNVLVIKQIKREGLIAVLLVCLISDVFLFI 54  
 Db 162 VFLSGLMGELKIGLLDFLLFGSLIAA VDPVAVLAVFEVH-----VNEVLFI 209  
 QY 55 AGTLGVLLSNAPIVL-----DIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIEETEP 109  
 Db 210 --VFGESLLNDAVTVLYNVFVSFTLGG-----DAVTGDCVKGVSPFV- 253  
 QY 110 TVPDDTFLGGSV-----ATDTRNRVRVSVSKQVWVKPMLMAIVLTWLNPN 159  
 Db 254 -----VSLGGTLGVIFAFLLSLVTRTKVRI---IEPGFVFI SYLSLTSEMLSLSA 305  
 QY + 160 YLDAVFVIG-----GVGAQYGDTRWTFAPAGAAFAASLIWFPPLVFGAAALSRLSS 210  
 Db 306 IL-AITFCGICCKYKVRANISEQSAITVRYTMKLSGAETIIFMFLGISAV-----D 357

QY 211 PKVWRWNVVAVVMTALAI 230  
 Db 358 FVITWTNTAFVILLTLVFISV 377  
 RESULT 13  
 US-09-097-053-11  
 ; Sequence 11, Application US/09097053  
 ; Patent No. 6392025  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Brant, Steven R.  
 ; APPLICANT: Yun, Chris C.H.  
 ; APPLICANT: Donowitz, Mark  
 ; APPLICANT: Tse, Chung-Ming  
 ; TITLE OF INVENTION: Cloning, Tissue Distribution, and  
 ; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,  
 ; TITLE OF INVENTION: NHE3.  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
 ; ADDRESSEE: Dunner  
 ; STREET: 1300 I Street, N.W., Suite 700  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005-3315  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/097,053  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/677,734  
 ; FILING DATE: 10-JUL-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fordis, Jean B.  
 ; REGISTRATION NUMBER: 32,984  
 ; REFERENCE/DOCKET NUMBER: 05387.0043-00000  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 408-4000  
 ; TELEFAX: (202) 408-4400  
 ; INFORMATION FOR SEQ ID NO: 11:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 831 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-09-097-053-11

Query Match 6.6%; Score 78.5; DB 4; Length 831;  
 Best Local Similarity 20.4%; Pred. No. 2.1;  
 Matches 53; Conservative 35; Mismatches 93; Indels 79; Gaps 11;  
 QY 6 IFITGL-----LLGASLLLSIGPQNVLVIKQIKREGLIAVLLVCLISDVFLFI 54  
 Db 162 VFLSGLMGELKIGLLDFLLFGSLIAA VDPVAVLAVFEVH-----VNEVLFI 209  
 QY 55 AGTLGVLLSNAPIVL-----DIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIEETEP 109  
 Db 210 --VFGESLLNDAVTVLYNVFVSFTLGG-----DAVTGDCVKGVSPFV- 253  
 QY 110 TVPDDTFLGGSV-----ATDTRNRVRVSVSKQVWVKPMLMAIVLTWLNPN 159  
 Db 254 -----VSLGGTLGVIFAFLLSLVTRTKVRI---IEPGFVFI SYLSLTSEMLSLSA 305  
 QY + 160 YLDAVFVIG-----GVGAQYGDTRWTFAPAGAAFAASLIWFPPLVFGAAALSRLSS 210



Db 360 -----NPEIWTWNTAFVLLTLVFISV 380

Search completed: April 24, 2003, 18:24:13  
Job time : 21 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 18:49:18 ; Search time 6117.94 Seconds  
(without alignments)  
11293.030 Million cell updates/sec

Title: US-09-105-117K-3

Perfect score: 2374

Sequence: 1 agataactcttggasgaaa.....gtaacaccttcagcaaatgg 2374

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	2374	100.0	2374	1	CGLYSEG X96471 C. glutamicu
C 2	2374	100.0	2374	6	A93933 Sequence 2
C 3	2374	100.0	333150	1	AP005277 Corynebac
C 4	2374	100.0	349980	6	AX127147 Sequence
C 5	993	41.8	993	6	AX063767 Sequence
C 6	993	41.8	993	6	AX244055 Sequence
C 7	870	36.6	870	6	AX123540 Sequence
C 8	822	34.6	822	6	AX063771 Sequence
C 9	822	34.6	822	6	AX244059 Sequence
C 10	798.8	33.6	1771	1	AB083133 Corynebac
C 11	708	29.8	708	6	AX123539 Sequence
C 12	627	26.4	627	6	AX123538 Sequence
C 13	485.2	20.4	993	6	AX067087 Sequence
C 14	170	7.2	345783	1	AP003001 Mesorhizo
C 15	164.4	6.9	8472	1	AE008076 Agrobacte
C 16	164.4	6.9	13051	1	AE009111 Agrobacte
C 17	159.4	6.7	15866	1	SC7H9
C 18	140.6	5.9	33285	1	SC5F8
C 19	135.6	5.7	306250	1	SME591788
C 20	122.8	5.2	14844	1	MBU34849
C 21	122.8	5.1	15239	6	I86263 Sequence 17
C 22	121.2	5.1	14869	1	AE007056 Mycobacte
C 23	121.2	5.1	38500	1	MTCY39
C 24	116.4	4.9	11120	1	RFA311775
C 25	113	4.8	36583	1	SCSH1
C 26	112.6	4.7	10871	1	AE011965
C 27	112.4	4.7	10336	1	AE005530 Xanthomon
C 28	112.2	4.7	12070	1	AB011413 Streptomy
C 29	111.2	4.7	266658	1	AP002563 Escherich
C 30	109.8	4.6	1525	1	PSE1GRA
C 31	108	4.5	11024	1	AE000382 Pseudomonas
C 32	108	4.5	141744	1	ECU28377 Escherich
C 33	107	4.5	110000	2	AC020884.0
C 34	105.2	4.4	22929	1	AE008808 Mus muscu
C 35	105.2	4.4	256050	1	AL627274 Salmonell
C 36	101	4.3	1200	6	AX065613 Sequence
C 37	100.8	4.2	5541	6	AR012071 Sequence
C 38	100.8	4.2	5541	6	AR025196 Sequence
C 39	100.8	4.2	5541	6	AR038653 Sequence
C 40	100.8	4.2	5541	6	AR068500 Sequence
C 41	100.8	4.2	5541	6	AX343066 Sequence
C 42	100.8	4.2	5541	6	AX347518 Sequence
C 43	100.8	4.2	5541	6	AX347541 Sequence
C 44	100.8	4.2	5541	6	AX347564 Sequence
C 45	93.2	3.9	988	6	AX431662 Sequence

## ALIGNMENTS

RESULT 1  
CGLYSEG/c  
LOCUS CGLYSEG 2374 bp DNA linear BCT 18-MAR-2001  
DEFINITION C. glutamicum lysE and lysG genes.  
ACCESSION X96471  
VERSION X96471.1 GI:1729753  
KEYWORDS lysE gene; lysG gene; Lysine export regulator protein; Lysine  
exporter protein; Lysine governor.  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum.  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 2374)  
AUTHORS Vrljic, M., Sahm, H. and Eggeling, L.



QY 1201 GAA CAA AAG ACGT CAG AAT TAA A C A C A C G A G A A G A A C C G C A A T G A G T C C T T C G C G C T T 1260  
Db 1174 GAA CAA AAG ACGT CAG AAT TAA A C A C A C G A G A A G A A C C G C A A T G A G T C C T T C G C G C T T 1115  
QY 1261 AAT T C C T T G T T A T C A C C A G T A C A T T T C G G G T C C G A T G A C A G T A A A G A C T G G G C C C 1320  
Db 1114 AAT T C C T T G T T A T C A C C A G T A C A T T T C G G G T C C G A T G A C A G T A A A G A C T G G G C C C 1055  
QY 1321 CAA A G C A G A C C T G T A A T G A A G T T T C A T G A T C A C C A T C G T G A C C T A T G A G T A C T T A 1380  
Db 1054 CAA A G C A G A C C T G T A A T G A A G T T C A T G A T C A C C A T C G T G A C C T A T G A A G T A C T T A 995  
QY 1381 A G T A A A T G A T T G T T C T T A A C A T G T T A A T A T A G C T T C A T G A A C C C A T T C A A C T G G A 1440  
Db 994 A G T A A A T G A T T G T T C T T A A C A T G T T A A T A T A G C T T C A T G A A C C C A T T C A A C T G G A 935  
QY 1441 C A C T T T G C T C A A T C A T T A T G A G G A G C A G T T C G A A G G C G C T C C T T A G C C T T T C A T 1500  
Db 934 C A C T T T G C T C A A T C A T T A T G A G G A G C A G T T C G A A G G C G C T C C T T A G C C T T T C A T 875  
QY 1501 T T C C C C T C G C G G T A G T C A G C C G T T A A A G C T C T C G A G C A T C A C G T G G T C G A G T T 1560  
Db 874 T T C C C C T C G C G G T A G T C A G C C G T T A A A G C T C T C G A G C A T C A C G T G G T C G A G T T 815  
QY 1561 G G T A T C G G C A C C A A C C G G C A A A G C A A C C G A A G C G G T G A A G T C C T T G T G C A A G C A G C 1620  
Db 814 G G T A T C G G C A C C A A C C G G C A A A G C A A C C G A A G C G G T G A A G T C C T T G T G C A A G C A G C 755  
QY 1621 G C G A A A T G G T G T G T C G A A G C A G A A C T A A A G C C A C T A T C G A G C C T T G C T G A 1680  
Db 754 G C G A A A T G G T G T G T C G A A G C A G A A C T A A A G C C A C T A T C G A G C C T T G C T G A 695  
QY 1681 A A T C C C G T A A C C A T C C C A T C A A G C A G A T T C G T A T C C A C A T G G T T C C T C C G G T T 1740  
Db 694 A A T C C C G T A A C C A T C C C A T C A A G C A G A T T C G T A T C C A C A T G G T T C C T C C G G T T 635  
QY 1741 C A A C G A G T A G T T T T G G G T G G A C A G C T C A C G C T G C G T T G G A A G A T G A A G C G C A 1800  
Db 634 C A A C G A G T A G T T T T G G G T G G A C A G C T C A C G C T G C G T T G G A A G A T G A A G C G C A 575  
QY 1801 C A C A T T A C C T T G T C G G G T G G A G A T G T T T T A G A G C G G T A A C C G T G A A G C T A A T C 1860  
Db 574 C A C A T T A C C T T G T C G G G T G G A G A T G T T T T A G A G C G G T A A C C G T G A A G C T A A T C 515  
QY 1861 C G T G G C G G A T G T G A G T A G A C T T G G A C C A T G C G C A C T T G C C A T T G C A C C C 1920  
Db 514 C G T G G C G G A T G T G A G T A G A C T T G G A C C A T G C G C A C T T G C C A T T G C A C C C 455  
QY 1921 C T C A T T C G G G A T C C C T A C A T G T T G A T G G A A A C T A G A T T G G G C T G C G A T G C C G T C T T 1980  
Db 454 C T C A T T C G G G A T C C C T A C A T G T T G A T G G A A A C T A G A T T G G G C T G C G A T G C C G T C T T 395  
QY 1981 A C G T T C G G T C C C A A G A T G T C T T A A G A C C G T A C T T G A G C G G G C G T C G A T G T C 2040  
Db 394 A C G T T C G G T C C C A A A G A T G T C T T A A G A C C G T A C T T G A G C G G G C G T C G A T G T C 335  
QY 2041 T G T G G G C G C A G C G G T A T C A T T G T C C G T C G G G A A G T T T G T G A G G C A A T T G 2100  
Db 334 T G T G G G C G C A G C G G T A T C A T T G T C C G T C G G G A A G T T T G T G A G G C A A T T G 275  
QY 2101 C C A G G C T T G T T G G G A C T T C T T C C G A A A C C A A G C T G C T C C C A T T G T A A A A G C A G G 2160  
Db 274 C C A G G C T T G T T G G G A C T T C T T C C G A A A C C C A A G C T G C T C C C A T T G T A A A A G C A G G 215  
QY 2161 A G A A G T A T C C T C T C A T G A T A C C A T T G A C A C A G A T A T T G G C A A C A G T G G G 2220  
Db 214 A G A A G T A T C C T C T C A T G A T A C C A T T G A C A C A G A T A T T G G C A A C A G T G G G 155  
QY 2221 C C T G G A A T C T A G A T C T A G C A G C C G T C G T T G A T G C A G C A A T C A G G 2280  
Db 154 C C T G G A A T C T A G A T C T A G C A G C C G T C G T T G A T G C A G C A A T C A G G 95

QY 2281 ATTGGCGCCTTACTTACTTCTGAAAAAGTTTCAGGGTTTTTCACTTCTTCCGCCGAGGAA 2340  
Db 94 ATTGGCGCCTTACTTACTTCTGAAAAAGTTTCAGGGTTTTTCACTTCTTCCGCCGAGGAA 35  
QY 2341 TTGGGCGAGGAGAGTAACACCTTTCAGCAAAATGG 2374  
Db 34 TTGGGCGAGGAGAGTAACACCTTTCAGCAAAATGG 1  
RESULT 2  
A93933/c A93933 Sequence 2 from Patent WO9723597. 2374 bp DNA linear PAT 22-JAN-2000  
LOCUS  
DEFINITION  
ACCESSION A93933  
VERSION A93933.1 GI:6742037  
KEYWORDS  
SOURCE  
ORGANISM  
Corynebacterium glutamicum.  
Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE  
1. (bases 1 to 2374)  
Vrlijc.M. and Eggeling, L.  
PROCESS FOR THE MICROBIAL PRODUCTION OF AMINO ACIDS BY BOOSTED  
ACTIVITY OF EXPORT CARRIERS  
JOURNAL Patent: WO 9723597-A 2 03-JUL-1997;  
KERNFORSCHUNGSANLAGE JUELICH (DE); VRLIJC MARINA (DE)  
FEATURES  
source  
1. .2374  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
BASE COUNT 526 a 640 c 648 g 560 t  
ORIGIN  
Query Match 100.0%; Score 2374; DB 6; Length 2374;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGATACTCCTTTGGAGAAACCATGTACCATTTGGTGACATTTGGTCTCGAAAGGC 60  
Db 2374 AGATACTCCTTTGGAGAAACCATGTACCATTTGGTGACATTTGGTCTCGAAAGGC 2315  
QY 61 TCTTTACGTGGGTATTCTTCTACGGTCCAGAGCTCACAGCGGCGGCTGAGTTTCA 120  
Db 2314 TCTTTACGTGGGTATTCTTCTACGGTCCAGAGCTCACAGCGGCGGCTGAGTTTCA 2255  
QY 121 GCGGAGGAGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCCATCAATTAATCGTTG 180  
Db 2254 GCGGAGGAGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCCATCAATTAATCGTTG 2195  
QY 181 GGTGGAGGAACCGGGCGATGAGGTGAGNACTTTGTTGAGTCAGCTGCCAACAATGGTCT 240  
Db 2194 GGTGGAGGAACCGGGCGATGAGGTGAGNACTTTGTTGAGTCAGCTGCCAACAATGGTCT 2135  
QY 241 TGGCGTCATTGCTTTCTACCACTTTCGCGAGGCTGCTTCACGGAACAATATCTCGATGG 300  
Db 2134 TGGCGTCATTGCTTTCTACCACTTTCGCGAGGCTGCTTCACGGAACAATATCTCGATGG 2075  
QY 301 AATTCAGAGGTTTCCCGCGCAGCGAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 360  
Db 2074 AATTCAGAGGTTTCCCGCGCAGCGAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 2015  
QY 361 GAACATATTGATATGGTCCGCAAGCTCAATGACATCGCCAGGAACCGGGCAGTCACT 420  
Db 2014 GAACATATTGATATGGTCCGCAAGCTCAATGACATCGCCAGGAACCGGGCAGTCACT 1955  
QY 421 TGGCAGATGGCGCTTGCATGGGTCTCGCGAGCAAGAGAGTACGGCGCGATACCGT 480  
Db 1954 TGGCAGATGGCGCTTGCATGGGTCTCGCGAGCAAGAGAGTACGGCGCGATACCGT 1895  
QY 481 GACCAAGTCATTTGTTGGTCTTCGTTCAGTTCAGAGCTGGCAACAGCCTTGAATTCAT 540  
Db 1894 GACCAAGTCATTTGTTGGTCTTCGTTCAGTTCAGAGCTGGCAACAGCCTTGAATTCAT 1835

QY 541 CAACAACCTGGAGCTTTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACGACGC 600  
DB 1834 CAACAACCTGGAGCTTTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACGACGC 1775  
QY 601 CGGCATCAACATTTGGGCGAAGCCACCGATTCACAAAACCCGGAAGAACTAAACCCATCAA 660  
DB 1774 CGGCATCAACATTTGGGCGAAGCCACCGATTCACAAAACCCGGAAGAACTAAACCCATCAA 1715  
QY 661 CATCAGTTTGAATGCCAATCGCGTCTATCACTACGACGACGATGATCCAGCGCCA 720  
DB 1714 CATCAGTTTGAATGCCAATCGCGTCTATCACTACGACGACGATGATCCAGCGCCA 1655  
QY 721 CACTTGGGGCTGAGACGCGGGGCTGACAAATGCTGTGCGCCGAAACCCACGACGCGGAA 780  
DB 1654 CACTTGGGGCTGAGACGCGGGGCTGACAAATGCTGTGCGCCGAAACCCACGACGCGGAA 1595  
QY 781 CCAGATCAGGCTTGGCGGAAACCGCGGCGGAAATCCACCGTCCGGTGTGCGCGTA 840  
DB 1594 CCAGATCAGGCTTGGCGGAAACCGCGGCGGAAATCCACCGTCCGGTGTGCGCGTA 1535  
QY 841 TTGCGGCGGACGCGCGGATAAACAACGCGGTCCAAATACGCAATTCGGGTTCACCA 900  
DB 1534 TTGCGGCGGACGCGCGGATAAACAACGCGGTCCAAATACGCAATTCGGGTTCACCA 1475  
QY 901 GGTACGACGATTTGCCATCAACATGAGGCTTTACGCAACCCGCTGCTTATCGAGCTCAC 960  
DB 1474 GGTACGACGATTTGCCATCAACATGAGGCTTTACGCAACCCGCTGCTTATCGAGCTCAC 1415  
QY 961 CTCACCGGACGCGGTGCGGCTGTCAGTGGCCACCGCGGAAACCCGCAAGCGGTGTC 1020  
DB 1414 CTCACCGGACGCGGTGCGGCTGTCAGTGGCCACCGCGGAAACCCGCAAGCGGTGTC 1355  
QY 1021 ATCGGGCAGGTTGGTCTGTTCTTCAATGATCTGTGGCGCTTCCACCTTTGTGTAT 1080  
DB 1354 ATCGGGCAGGTTGGTCTGTTCTTCAATGATCTGTGGCGCTTCCACCTTTGTGTAT 1295  
QY 1081 GCGCTTTTCTGCTGCTATGACGGAACCAATACAGGTAAAGCATGCGCCACCGCGCAT 1140  
DB 1294 GCGCTTTTCTGCTGCTATGACGGAACCAATACAGGTAAAGCATGCGCCACCGCGCAT 1235  
QY 1141 AATATCAGACGATCGGCGGCTTGGACAAAAGATCAACGCCCAAGGTGCGCGCAT 1200  
DB 1234 AATATCAGACGATCGGCGGCTTGGACAAAAGATCAACGCCCAAGGTGCGCGCAT 1175  
QY 1201 GAACAAAAGACGTCAGAAATTAACAACACGAGAAGAACCGCAATGAGTCTTTCGCGCTT 1260  
DB 1174 GAACAAAAGACGTCAGAAATTAACAACACGAGAAGAACCGCAATGAGTCTTTCGCGCTT 1115  
QY 1261 AATTCCTTTTATCACCAGTATCTTCGGTTCGATGGAAGATTAAGATGCGCC 1320  
DB 1114 AATTCCTTTTATCACCAGTATCTTCGGTTCGATGGAAGATTAAGATGCGCC 1055  
QY 1321 CAAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTACCTATGTAAGATCTTA 1380  
DB 1054 CAAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTACCTATGTAAGATCTTA 995  
QY 1381 AGTAAATATGATGGTCTTAAACATGTTTAAATATAGTCTTATGAACCCCAATCAACTGGA 1440  
DB 994 AGTAAATATGATGGTCTTAAACATGTTTAAATATAGTCTTATGAACCCCAATCAACTGGA 935  
QY 1441 CACTTTGCTCTCAATCATTTATGAGGACGCTTCGAAGGCGCTCTTAGCCCTTTCCAT 1500  
DB 934 CACTTTGCTCTCAATCATTTATGAGGACGCTTCGAAGGCGCTCTTAGCCCTTTCCAT 875  
QY 1501 TTCCCTCGGCGGTGAGTCAAGCGCTTAAAGCTCTCGAGCATCAGCTGGGTGAGTGT 1560  
DB 874 TTCCCTCGGCGGTGAGTCAAGCGCTTAAAGCTCTCGAGCATCAGCTGGGTGAGTGT 815  
QY 1561 GGTATCGGCAACCAACCGGCAAGCAACCGAAGCGGGTGAAGTCTTGTGCAAGCAGC 1620  
DB 814 GGTATCGGCAACCAACCGGCAAGCAACCGAAGCGGGTGAAGTCTTGTGCAAGCAGC 755

QY 1621 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGACGCGCTTGTCTGA 1680  
DB 754 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGACGCGCTTGTCTGA 695  
QY 1681 AATCCCGTTAAACCATCGCCATCAAGCGAGATTCGCTATCCACATGTTTCTTCCCGTGT 1740  
DB 694 AATCCCGTTAAACCATCGCCATCAAGCGAGATTCGCTATCCACATGTTTCTTCCCGTGT 635  
QY 1741 CAACGAGTAGCTTTTGGGGTGGAGCAACGCTACGCTGCGCTTGGAGATGAAGCGCA 1800  
DB 634 CAACGAGTAGCTTTTGGGGTGGAGCAACGCTACGCTGCGCTTGGAGATGAAGCGCA 575  
QY 1801 CACATTTCTTCTGCGCGGTGGAGATGTTTAAAGCGGTAAACCGTGAAGCTAATCC 1860  
DB 574 CACATTTCTTCTGCGCGGTGGAGATGTTTAAAGCGGTAAACCGTGAAGCTAATCC 515  
QY 1861 CGTGCGGAGTCTGAAGTAGTAGAATCTTGAACCATGCGCCACTTGGCCATTTGCAACCC 1920  
DB 514 CGTGCGGAGTCTGAAGTAGTAGAATCTTGAACCATGCGCCACTTGGCCATTTGCAACCC 455  
QY 1921 CTCTTGGCGGATGCTTACATGTTGATGGAAACTAGATTGGGCTGCGATGCCGTCTT 1980  
DB 454 CTCTTGGCGGATGCTTACATGTTGATGGAAACTAGATTGGGCTGCGATGCCGTCTT 395  
QY 1981 AGCTTCGCTCCCAAGATGCTGCTTCAAGACCGTCACTGACGCGCGCTCGATGGTCC 2040  
DB 394 AGCTTCGCTCCCAAGATGCTGCTTCAAGACCGTCACTGACGCGCGCTCGATGGTCC 335  
QY 2041 TGTGGGCGCGACGCGCTATCCATTTGCTCCGTCGCGGAAGGTTTGTGTGAGCAATTCG 2100  
DB 334 TGTGGGCGCGACGCGCTATCCATTTGCTCCGTCGCGGAAGGTTTGTGTGAGCAATTCG 275  
QY 2101 CCGAGGCTTGTGTGGGACCTTCTCCGAAACCAAGCTGCTCCCATGCTCTAAAAGCAG 2160  
DB 274 CCGAGGCTTGTGTGGGACCTTCTCCGAAACCAAGCTGCTCCCATGCTCTAAAAGCAG 215  
QY 2161 AGAAGTGATCTCTCGATGAGATACCCATTCACACCGATGATTTGCAAGATGGCG 2220  
DB 214 AGAAGTGATCTCTCGATGAGATACCCATTCACACCGATGATTTGCAAGATGGCG 155  
QY 2221 CTTGGAATCTAGATCTTAGCTAGACTCACAGACGCGCTGTTGATGACAGCAATCGAGG 2280  
DB 154 CTTGGAATCTAGATCTTAGCTAGACTCACAGACGCGCTGTTGATGACAGCAATCGAGG 95  
QY 2281 ATTGGGCTTGTGTGTTTCTGAAAAGTTTCAAGGTTTCTCTTCCCGCGCAGGAA 2340  
DB 94 ATTGGGCTTGTGTGTTTCTGAAAAGTTTCAAGGTTTCTCTTCCCGCGCAGGAA 35  
QY 2341 TTGGGCGGCGAGATTAACACCTTCAGCAATGG 2374  
DB 34 TTGGGCGGCGAGATTAACACCTTCAGCAATGG 1

RESULT 3  
AP005277LOCUS  
DEFINITIONACCESSION  
VERSION

## KEYWORDS

SOURCE  
ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

AP005277 333150 bp . DNA linear BCT 08-AUG-2002  
Corynebacterium glutamicum ATCC 13032 DNA, complete genome, section 4/10.

AP005277 BA000036  
AP005277.1 GI:21323710

Corynebacterium glutamicum ATCC 13032 (strain:ATCC 13032) DNA.  
Corynebacterium glutamicum ATCC 13032  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.

1 Nakagawa,S

Complete genomic sequence of Corynebacterium glutamicum ATCC 13032  
Unpublished  
2 (bases 1 to 333150)  
Nakagawa,S.  
Direct Submission  
Submitted (24-MAY-2002) Satoshi Nakagawa, Kyowa Hakko Kogyo Co.

Ltd., Tokyo Research Laboratories; 3-6-6, Asahi-machi, Machida,  
Tokyo 194-8533, Japan (E-mail: snakagawa@xanagen.com,  
Tel: 81-44-829-3031, Fax: 81-44-813-1651)  
This sequence is conducted by collaboration of Kyowa Hakko Kogyo  
Co. Ltd. And Kitasato University.

## FEATURES

```

source
    1..333150
    /organism="Corynebacterium glutamicum ATCC 13032"
    /strain="ATCC 13032"
    /db_xref="taxon:196627"
    /note="ATCC 13032"
    /complement(63..1292)
    /gene="Cg10944"
    /gene="Cg10944"
    /complement(63..1292)
    /codon_start=1
    /transl_table=11
    /product="Hypothetical membrane protein"
    /protein_id="BAB98337.1"
    /db_xref="GI:21323711"
    /translation="MTNPTERNARLLIWANGLQNGDOIIVAAKTVLPMLQAAGAPG
    FLALAVPIREAGSMPLPOAATGMLVRLQTSKVVYVIGSGQFVSALGIGVAALFLRG
    WALGITVIVLALSLFRSMCSASKDVQGVKISKRGGLVTRATVIGVWGLVAGL
    AIAIFGSGHFRVLAADVASSFSWLFASIVFAIEYAKPATPKNAPSNPWRRCI
    AALDKAFRRFVLRSMMLTALSTAFIVALAESGNSIDSLGFLIAGSLASWYGG
    RISGIWSDHSSKNMAGGALFGSIVLILVLSAFAPAOINTLVPLSPFLITLAHTA
    IRVARKTYVMDMAEGDQRTRYVADANTLMGVLLIIVGLSGFTAFIGNEAALLFLAAI
    GLAGTISARGLKEVSAG"
    1382..2917
    /gene="Cg10945"
    1382..2917
    /gene="Cg10945"
    /codon_start=1
    /transl_table=11
    /product="Putative multicopper oxidases"
    /protein_id="BAB98338.1"
    /db_xref="GI:21323712"
    /translation="MPENLRRTFFKGAVLAATVGAQVLVACSDDVRYGGEPRTL
    PTPADLGTREGSSVFALEAQTGESQILPDVTTKTWGNGTHLGPTLVVKKGDVHV
    DVNNIDEMTVHMGKMLPAIADGGPHSPIGQWSPPTVANDAAATLVKHPHTHG
    LTCGLHVRGLAGMLIVEDEATDKLDLPREYGVDDIPLVMDHRFLEDGSLDEEDLPDL
    GLIGDPTTANGIINAFDTRVRFRVLNGSNRFLNLAFSDDRTTFQVIASDSGLLD
    EPQDRTILAGPWEIIVLEPEGDVTLESVGFEDNYGVPDDEFPDFGMSDSFQL
    LTTGPSDDAAQAPALFGLVKSTEPDVIDATERPTIMNTFSINDIQMDQVADVVIL
    HDQPEYVINDNSDPHNFHVDARFKVLKPGTDVFLPNDGKMDKTVGLPGCATATL
    AVFEFGHPDPQWPMYHCHMLYHEDQGMGQFVIVEPGDEPAVLGSGTSSIDSAGG
    HAH"
    3063..3983
    /gene="Cg10946"
    3063..3983
    /gene="Cg10946"
    /note="PF00005:ABC transporter
    TIGR00960:3a0501s02: type II (General) Secretory Pathway
    (IISP) Family protein
    TIGR00968:3a0106s01: sulfate transport system permease
    protein
    TIGR01166:cbiO: cobalt transport protein ATP-binding
    subunit
    TIGR01184:ntrCD: nitrate transport ATP-binding subunits C
    and D
    TIGR01186:proV: glycine betaine/L-proline transport ATP
    binding subunit
    TIGR01187:potA: spermidine/putrescine ABC transporter
    ATP-binding subunit
    TIGR01188:draA: daunorubicin resistance ABC transporter
    ATP-binding subunit
    TIGR01189:ccmA: heme exporter protein CcmA"
    /codon_start=1
    /transl_table=11
    /product="ABC-type transporter, ATPase component"
    /protein_id="BAB98339.1"
    /db_xref="GI:21323713"
    /translation="MTTTPAIDVTDLVRTYGYDTAVKGLNFHFVQRGEVFLGTINGAG

```

```

KTSTLIEVIEGLSAPSSGTVRISGLDPVADRAILRELPGLQSGGLPSQLTVAETMDM
WHGTCYPRAIKQVLADVLLHRENVKVGSGGQRRLDLACALGDPSTILFLDEPT
TGLTDESRRHTWQLLDLKQGVYTWMLTHVLEAEFLCDRIATNNAGEIAVEGTLDE
LVARKSKIISFVLRGGQVELFVLSGAEIIRDNHVRVIAITTLQQTILSLTWAETGI
ALEGFAAKPATLSEVFMDSLENTSLQTA"
4003..4836
    /gene="Cg10947"
    4003..4836
    /gene="Cg10947"
    /codon_start=1
    /transl_table=11
    /product="ABC-type transporter, permease components"
    /protein_id="BAB98340.1"
    /db_xref="GI:21323714"
    /translation="MTTSHARGLOHATPERKTSFFKTSLFKAELQFRNKTLIFM
    ATVPVGVPLLLFLIGNGAAESANSFYFVNYTLLFVQFTVLSMATIRDERVLKR
    LRTGEARDIDIIIGACFFGALLTLFTVVIIPLLMVLGAPAPINLVPVFAVLIGLLI
    CSMALTSGFTRNAEAQMTSPVFMALMGGLGSIREFVGDSTIVADTLATVPFAAIS
    DLVQIGWAGATPADSVGCGVEANFAGIFQDMLIPLIGLAAWTAATAAANAAYMRWDSY
    R"
    4879..6138
    /gene="Cg10948"
    4879..6138
    /gene="Cg10948"
    /EC_number="2.7.3.-"
    /codon_start=1
    /transl_table=11
    /product="Two-component system, sensory transduction
    histidine kinases"
    /protein_id="BAB98341.1"
    /db_xref="GI:21323715"
    /translation="MNKDFWTAGTARFSGVSLASPVTAPLNSWPERLNLAKYTL
    YTRVSLQAIPIVLLSAYFLGIVANAGTLNPSFVLLGFSVILLITVTVLYEQSLNS
    HPRRSVQFFFTGLVNLVGVVSVLQIPGLNMSDNRATLALFTLTCVLLSLAYI
    PMNRYVWMLIAMSVALWMTSTTDDYLSALWVTPPLMAGTVRLSVMTVDVMKEVRS
    RELEASLRERFAQELHDTLQHLAAMSVKSELALAKRGDDRLLENRELQK
    LTRTSMSEMRDVVGYRTNLATEIEGAKSLADAHILSVIGTTSVQVSPARELCWA
    LVREATTILHLSDATDATLTLSTEVEMDNNGVKDIGRISGLSALRSRAESGMTL
    IVSREDDOFSVRMLINAPANTPAKEA"
    6139..6750
    /gene="Cg10949"
    6139..6750
    /gene="Cg10949"
    /note="PF00072:Response regulator receiver domain
    PF00196:Bacterial regulatory proteins, luxR family"
    /codon_start=1
    /transl_table=11
    /product="Two-component system, response regulators
    consisting of a CheY-like receiver domain and a HTH
    DNA-binding domain"
    /protein_id="BAB98342.1"
    /db_xref="GI:21323716"
    /translation="MISISADDEALIASLTLISLEPDLIDVPTAGSGRELITETWA
    DPSNRDVCVLDQLQGDIDTATRLMETTPDLAVLITSHARPROLKRALAAGVLG
    FLPTKSTADBEATARTVHAGRRVIDPELAATISAGESPLTNEEEVLELAGQGLSA
    BEIAVAHLAPGTRNVLVSQMTKVGQNRFEATRARELQWL"
    complement(6781..6987)
    /gene="Cg10950"
    complement(6781..6987)
    /gene="Cg10950"
    /codon_start=1
    /transl_table=11
    /product="Hypothetical protein"
    /protein_id="BAB98343.1"
    /db_xref="GI:21323717"
    /translation="MKHYWSMYFCAGPLNTPHTKVFMFMDIVSVNHWFEWSTNNWVN
    MTGLPRETFVPAFVGSTVLAVLSF"
    7048..8205
    /gene="Cg10951"
    7048..8205
    /gene="Cg10951"
    /codon_start=1
    /transl_table=11

```



```

/product="tRNA-processing ribonuclease BN"
/protein_id="BAB98344.1"
/db_xref="GI:21323718"
/translation="MTAWLCSPSPQVQDEYDMWNGVQVQEHLDATLIAADPHGNPEN
SGDRKERLNFQWKYALNRTVRDVRPDGLLDLAALITFPFSLISIAPAVLGYVITIP
LASDSTEILNLDVENVYVPEQSHVNGVIDSIAGSAAAGVGAVGVITVATLWTS
AVYRATRCANAVYGRSEGRTLIKRWAMLLFNALLLGLIIILVSWLNELVWGF
APIAEPLHLNLSFTLDRFPIWIWVRFPIVGVLMFVATLYWAPNAPKFRWL
SLGSPFLAIYVILLAGLNFYFTLFAAFSSYGAAGVSLAVFIALVFNICLIIGLKID
VEISRAKOLQAGMPADYSLVPPRSIERKVKQKQORQLMDQAAIREBN"
8672..10114
/gene="Cg10952"
CDS
8672..10114
/gene="Cg10952"
/notes="PF00005:ABC transporter
PF00664:ABC transporter transmembrane region."
/codon_start=1
/transl_table=1
/product="ABC-type multidrug/protein/lipid transport
system, ATPase component"
/protein_id="BAB98345.1"
/db_xref="GI:21323719"
/translation="MRTHTASSILRRMRIRQKGVAFGLFMWOLSEALVPIAI
GLIVDAVLTKDLRLVGLVAFVFLFVLSYREGSRALNRAVNFESHARVEVAD
HALKNDLPRLVPEVMSRSTADASTRIFGQITGVSAATGFLGAATYLLISWL
GLLVLPVPIISGVVALASKISKRSVTQOEKLAESGAQSDIMMLRVIKAGGERW
AVKTFKASQASAAVADVASGKAVIGELSIAVNLAAVLLAGWRVTTGELPGQ
LIAVPGAVLSPRIILNSINASAIHGAERVANFLNLDSEQAQYESSSTINDGE
LIVVPSTLPHGNDILATPHAADIFEGTLRSNTSMHEDNVPIDPOVIRASGLTDI
IEVDGDAFVRDGTNLSGQROVALRALHADAEVLVMDPSTSAVDSVTEVSLAQ
IKQLRAGKTTIVSSSPAPFNLDRAVISHV"
10107..11843
/gene="Cg10953"
CDS
10107..11843
/gene="Cg10953"
/notes="PF00005:ABC transporter

```

Query Match 100.0%; Score 2374; DB 1; Length 333150;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 AGATACCTCTTTGAAGAAACCATGTACGATTCGCGACATTTGTCGCTCGAAGGC 60
DB 327645 AGATACCTCTTTGAAGAAACCATGTACGATTCGCGACATTTGTCGCTCGAAGGC 327704
QY 61 TCTTTACGTGGTATTTCTTCTACGCTCCAGAGCTCACAGCGGAGCGGCTCAGTTTCAT 120
DB 327705 TCTTTACGTGGTATTTCTTCTACGCTCCAGAGCTCACAGCGGAGCGGCTCAGTTTCAT 327764
QY 121 GCGCGAGAGGGCTGCGCGCTCTGATTCATCAGCAGCTATTCATATTAATCGTTG 180
DB 327765 GCGCGAGAGGGCTGCGCGCTCTGATTCATCAGCAGCTATTCATATTAATCGTTG 327824
QY 181 GGTGGAGAACCGGGCGATGACGCTGAGAACTTGTTCAGTCAGCTCCCAACATGCTCT 240
DB 327825 GGTGGAGAACCGGGCGATGACGCTGAGAACTTGTTCAGTCAGCTCCCAACATGCTCT 327884
QY 241 TGGCGTCATTTCTTCTCACCACCTTGGCGAGGGCTCTGTCACGGACAAATATCTCGATGG 300
DB 327885 TGGCGTCATTTCTTCTCACCACCTTGGCGAGGGCTCTGTCACGGACAAATATCTCGATGG 327944
QY 301 AATTCCAGAGGGTTCGCGCGCAGCGAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 360
DB 327945 AATTCCAGAGGGTTCGCGCGCAGCGAGGTAAGTCCCTGTCTGAGGGCATGTTGAACGT 328004
QY 361 GAACAAATATTGATATGTTCGCAAGCTCAATGACATCGCCAGGAAACGCGGCGAGTCACT 420
DB 328005 GAACAAATATTGATATGTTCGCAAGCTCAATGACATCGCCAGGAAACGCGGCGAGTCACT 328064
QY 421 TGCGCAGATGCGCTTGCGATGGGTGCTGCGCGACGAAGGAGTACGCGCGGATACCGT 480
DB 328065 TGCGCAGATGCGCTTGCGATGGGTGCTGCGCGACGAAGGAGTACGCGCGGATACCGT 328124

```

```

QY 481 GACCAGTCATTGATTGTTGGTCTTGTGTCAGTTCAGCAGCTGGCAACAGCCTTGATTCACT 540
DB 328125 GACCAGTCATTGATTGTTGGTCTTGTGTCAGTTCAGCAGCTGGCAACAGCCTTGATTCACT 328184
QY 541 CAACAACTTTGGAGTTTCTTGACCGCGAGTTGGAGGCGATCGATGAGATTTCCTCCAGCGC 600
DB 328185 CAACAACTTTGGAGTTTCTTGACCGCGAGTTGGAGGCGATCGATGAGATTTCCTCCAGCGC 328244
QY* 601 CGGCATCAACATTTGGCGGAAGCCACCGATTCCAAATCCAAACCCCGGAAACCTAACCCATCAA 660
DB 328245 CGGCATCAACATTTGGCGGAAGCCACCGATTCCAAATCCAAACCCCGGAAACCTAACCCATCAA 328304
QY 661 CATCAGTTTGTATGGCCATGCGGTTCATCAACTGCCACGACGAGCTTGATCCAGCGCCA 720
DB 328305 CATCAGTTTGTATGGCCATGCGGTTCATCAACTGCCACGACGAGCTTGATCCAGCGCCA 328364
QY 721 CACCTTGGGGCTGGACAGCGGGCGGTGACAAATGCTGTCGCCGGAACCCACACGCGGGAA 780
DB 328365 CACCTTGGGGCTGGACAGCGGGCGGTGACAAATGCTGTCGCCGGAACCCACACGCGGGAA 328424
QY 781 CCAGATCAGGCTTGGCGGAACCGCGCCAGCGGCAAAATCCACCGTTCGGTGTGCGCGTA 840
DB 328425 CCAGATCAGGCTTGGCGGAACCGCGCCAGCGGCAAAATCCACCGTTCGGTGTGCGCGTA 328484
QY 841 TTGCGCGCGACGCGCGCGCATTAACACAAACCGGTCCAAATAGGCATTCGGGTTCACACCA 900
DB 328485 TTGCGCGCGCGACGCGCGCGCATTAACACAAACCGGTCCAAATAGGCATTCGGGTTCACACCA 328544
QY 901 GGTTCACGAGATTGCCATCAACATGGGCTTTACCCAAACCCGCTGCTTATCCAGCGCTCAC 960
DB 328545 GGTTCACGAGATTGCCATCAACATGGGCTTTACCCAAACCCGCTGCTTATCCAGCGCTCAC 328604
QY 961 CTCACCCCGCACCCCGTTCGCGGTGTGTCGCGCATTCAGTGGCCACCGCGCAACCCCAAGGCGTGC 1020
DB 328605 CTCACCCCGCACCCCGTTCGCGGTGTGTCGCGCATTCAGTGGCCACCGCGCAACCCCAAGGCGTGC 328664
QY 1021 ATCGGCGACGGTTGGTCTGTTTCTTCAATGATCTGTGGCGCTTCACCTTGTGTTGTCAT 1080
DB 328665 ATCGGCGACGGTTGGTCTGTTTCTTCAATGATCTGTGGCGCTTCACCTTGTGTTGTCAT 328724
QY 1081 GGGCTCTTTGCTGCCATGACCGGCAACCATTAACAGGTAAAGCATGCCACCCAGCGCAT 1140
DB 328725 GGGCTCTTTGCTGCCATGACCGGCAACCATTAACAGGTAAAGCATGCCACCCAGCGCAT 328784
QY 1141 AATATCGACGACGATCGCGCGCATTTGACAAAGATCAACGCCCAAGTGCCTGCGCGAT 1200
DB 328785 AATATCGACGACGATCGCGCGCATTTGACAAAGATCAACGCCCAAGTGCCTGCGCGAT 328844
QY 1201 GAACAAAAGAGCTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTCGCGCTT 1260
DB 328845 GAACAAAAGAGCTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTCGCGCTT 328904
QY 1261 AATTCTCTGTTTAAATCACCAGTACATTCGCGGTCCGATGGACAGTAAAGACTTGGCCCC 1320
DB 328905 AATTCTCTGTTTAAATCACCAGTACATTCGCGGTCCGATGGACAGTAAAGACTTGGCCCC 328964
QY 1321 CAAAACGACGCTGTAAATGAAGATTTCATGATCACCATCGTGACCTATGGAGTACTTA 1380
DB 328965 CAAAACGACGCTGTAAATGAAGATTTCATGATCACCATCGTGACCTATGGAGTACTTA 329024
QY 1381 AGTAAATGATTTGTTCTTAAACATGTTTAAATATAGCTTCATGAACCCCATTCACCTGGA 1440
DB 329025 AGTAAATGATTTGTTCTTAAACATGTTTAAATATAGCTTCATGAACCCCATTCACCTGGA 329084
QY 1441 CACTTTGCTCTCAATCATTTGATGAAGCGAGCTTCGAAGCGGCTCTTAGCCCTTTCCAT 1500
DB 329085 CACTTTGCTCTCAATCATTTGATGAAGCGAGCTTCGAAGCGGCTCTTAGCCCTTTCCAT 329144
QY 1501 TTCCTCCCTCGCGGGTGAAGTCAAGCGGTTAAAGCTCTCGAGCATCAAGTGGGTGAGTGT 1560
DB 329145 TTCCTCCCTCGCGGGTGAAGTCAAGCGGTTAAAGCTCTCGAGCATCAAGTGGGTGAGTGT 329204
QY 1561 GGTATCGCGCACCCACCGGCCCAAGCAACCGGAGGAGTGAAGTCTTGTGCAAGCAGC 1620

```

Db 329205 GGTATCGCGACCCAAACCGGCAAGCAACCGAGCGGTGAAGTCTTGTGCAAGCAGC 329264  
QY 1621 GCGGAAATGCTGCTGCAAGCAGAACTTAAGCGCACTATCTGAGCCCTTCTGA 1680  
Db 329265 GCGGAAATGCTGCTGCAAGCAGAACTTAAGCGCACTATCTGAGCCCTTCTGA 329324  
QY 1681 AATCCCGTTAAACCATCGCCATCAACGAGATTTGCTATCACATGGTTCTCCCGTGT 1740  
Db 329325 AATCCCGTTAAACCATCGCCATCAACGAGATTTGCTATCACATGGTTCTCCCGTGT 329384  
QY 1741 CAACGAGTGTCTTTGGGTGAGCAACGCTCACGCTGCTTGGAGATGAAGCGCA 1800  
Db 329385 CAACGAGTGTCTTTGGGTGAGCAACGCTCACGCTGCTTGGAGATGAAGCGCA 329444  
QY 1801 CACATATCTTCTGCGCGGTGAGATGTTTGAAGAGCGGTAAACCGCTGAAGCTAATCC 1860  
Db 329445 CACATATCTTCTGCGCGGTGAGATGTTTGAAGAGCGGTAAACCGCTGAAGCTAATCC 329504  
QY 1861 CGTGGCGGATGTGAAGTAGTAGAACTTGAACCATCGCCACTTGGCCATTTGCAACCCC 1920  
Db 329505 CGTGGCGGATGTGAAGTAGTAGAACTTGAACCATCGCCACTTGGCCATTTGCAACCCC 329564  
QY 1921 CTCAATTGCGGATGCTTACATGTTGATGGAAGAACTAGATGGGCTCGCATGCCGCTTT 1980  
Db 329565 CTCAATTGCGGATGCTTACATGTTGATGGAAGAACTAGATGGGCTCGCATGCCGCTTT 329624  
QY 1981 ACGCTTCGCTCCCAAGATGCTTCAAGACCGTGACCTGAGCGGGCGCTCGATGGTCC 2040  
Db 329625 ACGCTTCGCTCCCAAGATGCTTCAAGACCGTGACCTGAGCGGGCGCTCGATGGTCC 329684  
QY 2041 TGTGGGCGCAGGCGCGTATCCATTTGCTCCGTCGGCGGAAGTTTTGGTCGAGGCAATTCG 2100  
Db 329685 TGTGGGCGCAGGCGCGTATCCATTTGCTCCGTCGGCGGAAGTTTTGGTCGAGGCAATTCG 329744  
QY 2101 CCGAGGCTTGGTGGGGAATCTTCTCCGAAACCCAAAGTCTCCCATGCTTAAAGCAGG 2160  
Db 329745 CCGAGGCTTGGTGGGGAATCTTCTCCGAAACCCAAAGTCTCCCATGCTTAAAGCAGG 329804  
QY 2161 AGAAGTATCTCTCGATGATACCATTTGACACACCATGATTTGGCAAGTGGCG 2220  
Db 329805 AGAAGTATCTCTCGATGATACCATTTGACACACCATGATTTGGCAAGTGGCG 329864  
QY 2221 CTTGGAATCTAGATCTCTAGTAGACTCACAGACCGCTGTTGATGACAGCAATTCAGGG 2280  
Db 329865 CTTGGAATCTAGATCTCTAGTAGACTCACAGACCGCTGTTGATGACAGCAATTCAGGG 329924  
QY 2281 ATTGCGGCTTGTACTTCTGAAAAGTTTCAAGGTTTTCCTTCTTCCCGCCAGGAA 2340  
Db 329925 ATTGCGGCTTGTACTTCTGAAAAGTTTCAAGGTTTTCCTTCTTCCCGCCAGGAA 329984  
QY 2341 TTGGGCGCAGGACAGTAAACCTTCAGCAATGG 2374  
Db 329985 TTGGGCGCAGGACAGTAAACCTTCAGCAATGG 330018

RESULT 4  
AX127147  
LOCUS AX127147 349980 bp DNA linear PAT 11-MAY-2001  
DEFINITION Sequence 7063 from Patent EP1108790.  
ACCESSION AX127147 AX114121  
VERSION AX127147.1 GI:14041135  
KEYWORDS Corynebacterium glutamicum.  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 349980)  
AUTHORS Nakagawa, S., Mizoguchi, H., Ando, S., Hayashi, M., Ochiai, K.,  
Yokoi, H., Tateishi, N., Senoh, A., Ikeda, M. and Ozaki, A.  
TITLE Novel polynucleotides  
JOURNAL Patent: EP 1108790-A 7063 20-JUN-2001;

QYOWA HAKKO KOGYO CO., LTD. (JP)  
Location/Qualifiers  
1. 349980  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
/note="Seq 1 to long (3,309,400) split in 11, seq 7063  
1.200.001 1.549.980"  
BASE COUNT 80727 a 91049 c 97618 g 80586 t  
ORIGIN  
Query Match 100.0%; Score 2374; DB 6; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGATACCTCTTTGGAGAAACCATGTACGCTTGGGTGACATTTGGCTCTGGAAAGCG 60  
Db 127595 AGATACCTCTTTGGAGAAACCATGTACGCTTGGGTGACATTTGGCTCTGGAAAGCG 127654  
QY 61 TCTTTACGTGGGTATTTCTTCTTACGCTTCCAGAGCTCACAGCGAGCGGCTGAGTTTCA 120  
Db 127655 TCTTTACGTGGGTATTTCTTCTTACGCTTCCAGAGCTCACAGCGAGCGGCTGAGTTTCA 127714  
QY 121 GCGGAGGAGGGCTGCGCTTCTGATTCATCAGCAAGCTATTCATTAATTCGTTG 180  
Db 127715 GCGGAGGAGGGCTGCGCTTCTGATTCATCAGCAAGCTATTCATTAATTCGTTG 127774  
QY 181 GGTGGAGGAACCGGCGGATGACGTTGAGAACTTTGAGTCACTGCGCAACATGGTCT 240  
Db 127775 GGTGGAGGAACCGGCGGATGACGTTGAGAACTTTGAGTCACTGCGCAACATGGTCT 127834  
QY 241 TGGCGTCAATGCTTCTTCTACCACTTTCGCGAGGCTTCAAGTCTCTGAGGGCATGTTGAACGT 360  
Db 127835 TGGCGTCAATGCTTCTTCTTACCACTTTCGCGAGGCTTCAAGTCTCTGAGGGCATGTTGAACGT 127894  
QY 301 AATTCAGAGGGTTCCTCGCGCCAGCGAGGCTTAAAGTCTCTGAGGGCATGTTGAACGT 360  
Db 127895 AATTCAGAGGGTTCCTCGCGCCAGCGAGGCTTAAAGTCTCTGAGGGCATGTTGAACGT 127954  
QY 361 GAACATATGATATGCTCGCAAGCTTAATGACATGCGCCAGGAACGCGGGGAGTCACT 420  
Db 127955 GAACATATGATATGCTCGCAAGCTTAATGACATGCGCCAGGAACGCGGGGAGTCACT 128014  
QY 421 TGGCGAGATGGCGCTTGCATGCTGCTGCGAGCAAGGAGAGTACGGCGCGGATACCT 480  
Db 128015 TGGCGAGATGGCGCTTGCATGCTGCTGCGAGCAAGGAGAGTACGGCGCGGATACCT 128074  
QY 481 GACAGTGCATTTGATGTTGCTTCTGATGAGCAGCTGGAACAACAGCTTGAATTCAT 540  
Db 128075 GACAGTGCATTTGATGTTGCTTCTGATGAGCAGCTGGAACAACAGCTTGAATTCAT 128134  
QY 541 CAACACTTGGAGTTTCTGACGCGGAGTTGAGGGGATCGATGAGATTTCCACGACGC 600  
Db 128135 CAACACTTGGAGTTTCTGACGCGGAGTTGAGGGGATCGATGAGATTTCCACGACGC 128194  
QY 601 CGGCACTCAACATTTGGCGGAAGGCCACCATGTTCCAAAACCCCGAAATCAACCATCAA 660  
Db 128195 CGGCACTCAACATTTGGCGGAAGGCCACCATGTTCCAAAACCCCGAAATCAACCATCAA 128254  
QY 661 CATCAGTTTGTATGGCAATGCGGTTCATCAAACTGCGCAGCAGCTTGTGATCCAGCGCA 720  
Db 128255 CATCAGTTTGTATGGCAATGCGGTTCATCAAACTGCGCAGCAGCTTGTGATCCAGCGCA 128314  
QY 721 CACTTGGGGCTGGAACAGCGGCGGTGACAAATGCTGCTGCGCGCAAAACCCACAGCGGAA 780  
Db 128315 CACTTGGGGCTGGAACAGCGGCGGTGACAAATGCTGCTGCGCGCAAAACCCACAGCGGAA 128374  
QY 781 CCAGATCAGCTTTCGCGGAACCGCGCGGCAAAATCCACCTTCCGGTTCGCCCTGA 840  
Db 128375 CCAGATCAGCTTTCGCGGAACCGCGCGGCAAAATCCACCTTCCGGTTCGCCCTGA 128434  
QY 841 TTGCGCGCGCAGCGCGCGCGGATAAACCAAAACCGGTTCGAGATTCGCGGTTCAACCA 900  
Db 128435 TTGCGCGCGCAGCGCGCGCGGATAAACCAAAACCGGTTCGAGATTCGCGGTTCAACCA 128494

```
QY 901 GGTGAGCAGATGCTCCATCAACATGGGCTTTACCAAAACCGCGTCTTATCGAGCTCAC 960
Db 128495 GGTGAGCAGATGCTCCATCAACATGGGCTTTACCAAAACCGCGTCTTATCGAGCTCAC 128554
QY 961 CTCACCCGACACCGGTTGGCGGTGTGAGTGGCCACCGCGCAACCGCCCAAGGCGGTGTC 1020
Db 128555 CTCACCCGACACCGGTTGGCGGTGTGAGTGGCCACCGCGCAACCGCCCAAGGCGGTGTC 128614
QY 1021 ATCGGGCAGGTTGGTTCCTGTTCTTCAATGATCTGTGGCGCTTCCACCTTGTGTTCAT 1080
Db 128615 ATCGGGCAGGTTGGTTCCTGTTCTTCAATGATCTGTGGCGCTTCCACCTTGTGTTCAT 128674
QY 1081 GGCGTCTTTCGCTCCATGACGGCAACCATTAACAGTAAGCGATGCGCCACCGCGCAT 1140
Db 128675 GGCGTCTTTCGCTCCATGACGGCAACCATTAACAGTAAGCGATGCGCCACCGCGCAT 128734
QY 1141 AATATCGAGCAGATCGCGCGGCTTGGCAAAAGATCAACGCGCAAGGTGCGCGCAT 1200
Db 128735 AATATCGAGCAGATCGCGCGGCTTGGCAAAAGATCAACGCGCAAGGTGCGCGCAT 128794
QY 1201 GAACAAAGAAGCGTCAGAAATTAACACACGAGAAACCGCAATGATGCTTTCGCGCTT 1260
Db 128795 GAACAAAGAAGCGTCAGAAATTAACACACGAGAAACCGCAATGATGCTTTCGCGCTT 128854
QY 1261 AATTCCTTGTATTAACACAGTACATTCGCGGTCCGATGAGCAAGTAAGACATGCGCCC 1320
Db 128855 AATTCCTTGTATTAACACAGTACATTCGCGGTCCGATGAGCAAGTAAGACATGCGCCC 128914
QY 1321 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGAGTACTTA 1380
Db 128915 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGAGTACTTA 128974
QY 1381 AGTAAATGATGTTGTTCTTAACATGTTTAAATAGTCTTCATGAACCCCAATCAACTGGA 1440
Db 128975 AGTAAATGATGTTGTTCTTAACATGTTTAAATAGTCTTCATGAACCCCAATCAACTGGA 129034
QY 1441 CACTTCCTCTCATCATGATGAGGAGCTTCGAGGCGGCTTCAGCCCTTTCCAT 1500
Db 129035 CACTTCCTCTCATCATGATGAGGAGCTTCGAGGCGGCTTCAGCCCTTTCCAT 129094
QY 1501 TTCCCTCTCGCGGTGAGTCAGCGGCTTAAAGCTCTCGAGCATCAGCTGGGTGAGTGT 1560
Db 129095 TTCCCTCTCGCGGTGAGTCAGCGGCTTAAAGCTCTCGAGCATCAGCTGGGTGAGTGT 129154
QY 1561 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGCTGAAGTCTTGTGCAAGCAGC 1620
Db 129155 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGCTGAAGTCTTGTGCAAGCAGC 129214
QY 1621 GCGGAAATGTTGTTGTCGACGAGAACTAAAGCCCACTATCTGAGCGCTTGTGTA 1680
Db 129215 GCGGAAATGTTGTTGTCGACGAGAACTAAAGCCCACTATCTGAGCGCTTGTGTA 129274
QY 1681 AATCCCGTTAAACATCGCCATCAACGAGATTCGCTATCCACATGTTTCTCCCGGT 1740
Db 129275 AATCCCGTTAAACATCGCCATCAACGAGATTCGCTATCCACATGTTTCTCCCGGT 129334
QY 1741 CAACAGGTAGTCTTTGGGTGGAGCAACGCTCACGCTGCGCTTGAAGATGAAGCGCA 1800
Db 129335 CAACAGGTAGTCTTTGGGTGGAGCAACGCTCACGCTGCGCTTGAAGATGAAGCGCA 129394
QY 1801 CACATTAATCTTGTGCGGCTGAGATGTTTGAAGAGCGGTAAACCGGTGAAGCTATCC 1860
Db 129395 CACATTAATCTTGTGCGGCTGAGATGTTTGAAGAGCGGTAAACCGGTGAAGCTATCC 129454
QY 1861 CGTGGCGGATGTGAAGTAGTAACCTTGAACCATCGCCACCTTGGCCATGCAACCCC 1920
Db 129455 CGTGGCGGATGTGAAGTAGTAACCTTGAACCATCGCCACCTTGGCCATGCAACCCC 129514
QY 1921 CTCATTCGGGATGCTTACATGTTGTGATGGAAGCTAGATTTGGGCTCGCATGCGCGCTT 1980
Db 129515 CTCATTCGGGATGCTTACATGTTGTGATGGAAGCTAGATTTGGGCTCGCATGCGCGCTT 129574
```

```
QY 1981 AGCTTCGGTCCCAAGATGCTCTCAAGACCGTGACCTGGAGCGCGCTCGATGGTCC 2040
Db 129575 AGCTTCGGTCCCAAGATGCTCTCAAGACCGTGACCTGGAGCGCGCTCGATGGTCC 129634
QY 2041 TGTGGGCGCGAGCGCGGTATCCATTTGTCGCGCGGAAGGTTTTTGTGTAGGCAATTCG 2100
Db 129635 TGTGGGCGCGAGCGCGGTATCCATTTGTCGCGCGGAAGGTTTTTGTGTAGGCAATTCG 129694
QY 2101 CGAGGCGCTTGTGTGGGAGCTTCTTCCGAAACCAAGCTGCTCCCATCTCTAAAGCAGG 2160
Db 129695 CGAGGCGCTTGTGTGGGAGCTTCTTCCGAAACCAAGCTGCTCCCATCTCTAAAGCAGG 129754
QY 2161 AGAAGTGATCTCTCGATGAGATACCCATTCACACACCGATGATTTGGCAACGATGGCG 2220
Db 129755 AGAAGTGATCTCTCGATGAGATACCCATTCACACACCGATGATTTGGCAACGATGGCG 129814
QY 2221 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCGCTGTTGATGACAGCAATCGAGG 2280
Db 129815 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCGCTGTTGATGACAGCAATCGAGG 129874
QY 2281 ATTGGCGCTTGTAGTACTTCTGAAAGGTTTCAGGGTTTTTCACTTCTCGCCGCGAGGAA 2340
Db 129875 ATTGGCGCTTGTAGTACTTCTGAAAGGTTTCAGGGTTTTTCACTTCTCGCCGCGAGGAA 129934
QY 2341 TTGGGCGCGCAGAGTAACACCTTCAGCAATGG 2374
Db 129935 TTGGGCGCGCAGAGTAACACCTTCAGCAATGG 129968
```

RESULT 5  
AX063767  
LOCUS AX063767 993 bp DNA linear PAT 24-JAN-2001  
DEFINITION Sequence 49 from Patent WO0100843.  
ACCESSION AX063767  
VERSION AX063767.1 GI:12541479  
KEYWORDS  
SOURCE Corynebacterium glutamicum.  
ORGANISM Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE 1 (bases 1 to 993)  
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O. and Haberhauer, G.  
TITLE corynebacterium glutamicum genes encoding metabolic pathway  
JOURNAL Patent: WO 0100843-A 49 04-JAN-2001;  
BASF AKTIENGESSELLSCHAFT (DE)  
FEATURES  
source  
1. 993  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
101..973  
/note="BXAO1393"  
/codon\_start=1  
/transl\_table=11  
/db\_xref="GI:12541480"  
/translation="MNPQLDITLLSIIDEGSFEGLASLISPSVSQVKALEHHVG  
RVLVSRTPAKATEAGEVLVQAKRMVLQAEATKQALSGRLAEIPLTINAINSLSTW  
PPVPNEVASWGATLRLRLDEAHTLSLLRRGDLVGLAVTREANVAGCEVVELGTM  
HLAINTPSLRDAYMDVKLDWAMPVLRFGPKVDLQDRDLQGRVDGPPGRRRVSIVPS  
AEGFGEAIRGLGWLGPETQAPMLKAGEVILLDFIDITPMYQWRWRLRSRLAR  
TDVVDAAEGLRP"

BASE COUNT 233 a 256 c 268 g 236 t  
ORIGIN  
Query Match 41.8%; Score 993; DB 6; Length 993;  
Best Local Similarity 100.0%; Pred. No. 2.4e-233;  
Matches 993; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1321 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGAGTACTTA 1380  
Db 1 CAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGAGTACTTA 60

Qy 1381 AGTAAATGATGGTCTTAAATAGCTTTAATATAGCTTCAATGAAACCCATTCAACTGGA 1440  
Db 61 AGTAAATGATGGTCTTAAATAGCTTTAATATAGCTTCAATGAAACCCATTCAACTGGA 120  
Qy 1441 CACTTTGCTCTCAATCATATGATGAGGAGCTTGAAGCGGCGCTCTTTAGCCCTTTCCAT 1500  
Db 121 CACTTTGCTCTCAATCATATGATGAGGAGCTTGAAGCGGCGCTCTTTAGCCCTTTCCAT 180  
Qy 1501 TTCCCGCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACGTTGGGTGAGTGT 1560  
Db 181 TTCCCGCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACGTTGGGTGAGTGT 240  
Qy 1561 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGCTGAAGTCTTGTGCAAGCAGC 1620  
Db 241 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGCTGAAGTCTTGTGCAAGCAGC 300  
Qy 1621 CGGAAATGTTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGAGCGCTTGCTGA 1680  
Db 301 CGGAAATGTTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGAGCGCTTGCTGA 360  
Qy 1681 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCACATGTTTCTCCCGTGT 1740  
Db 361 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCACATGTTTCTCCCGTGT 420  
Qy 1741 CAACGAGTAGCTTCTTTGGGTGGAGCAACGCTCAACGCTGCGCTTGAAGATGAAGCGCA 1800  
Db 421 CAACGAGTAGCTTCTTTGGGTGGAGCAACGCTCAACGCTGCGCTTGAAGATGAAGCGCA 480  
Qy 1801 CACATTATCTTGTGCGGCTGGAGATGTTTTAGAGCGGTAAACCGTGAAGCTAATCC 1860  
Db 481 CACATTATCTTGTGCGGCTGGAGATGTTTTAGAGCGGTAAACCGTGAAGCTAATCC 540  
Qy 1861 CGTGGCGATGTCAGTAGTAGTGAAGCTTGAACCATGCGCACCTGCGCATTTGCCATTCGAAACCC 1920  
Db 541 CGTGGCGATGTCAGTAGTAGTGAAGCTTGAAGCATGCGCACCTGCGCATTTGCCATTCGAAACCC 600  
Qy 1921 CTCAATCGGATGCTACATGTTGATGGAAACTAGATTGGGCTGCGATGCCCGCTT 1980  
Db 601 CTCAATCGGATGCTACATGTTGATGGAAACTAGATTGGGCTGCGATGCCCGCTT 660  
Qy 1981 ACGCTTGGTCCCAAGATGCTTCAAGACCGTGACCTGACCGGCGGCTGATGTC 2040  
Db 661 ACGCTTGGTCCCAAGATGCTTCAAGACCGTGACCTGACCGGCGGCTGATGTC 720  
Qy 2041 TGTGGGCGCAGGCGGATGCTATGCTTCCGTCGCGGAGGTTTGTGAGGCAATTCG 2100  
Db 721 TGTGGGCGCAGGCGGATGCTATGCTTCCGTCGCGGAGGTTTGTGAGGCAATTCG 780  
Qy 2101 CCGAGGCTTGGTGGGACTTCTTCCGAAACCAAGCTGCTCCCATGCTTAAAGCAGG 2160  
Db 781 CCGAGGCTTGGTGGGACTTCTTCCGAAACCAAGCTGCTCCCATGCTTAAAGCAGG 840  
Qy 2161 AGAAGTATCTCTCGATGATGATCCCATGACACACCGATGATTTGGCAACGATGGCG 2220  
Db 841 AGAAGTATCTCTCGATGATGATCCCATGACACACCGATGATTTGGCAACGATGGCG 900  
Qy 2221 CCTGAACTAGATCTCTAGCTAGACTACAGACGCGCTGTTGATGCAATTCGAGG 2280  
Db 901 CCTGAACTAGATCTCTAGCTAGACTACAGACGCGCTGTTGATGCAATTCGAGG 960  
Qy 2281 ATTGCGGCTTAGTTACTTCTGAAAGGTTTCAG 2313  
Db 961 ATTGCGGCTTAGTTACTTCTGAAAGGTTTCAG 993

RESULT 6  
AX244055  
LOCUS AX244055  
DEFINITION Sequence 47 from Patent WO0166573.  
ACCESSION AX244055  
VERSION AX244055.1 GI:15859119  
KEYWORDS linear PAT 29-SEP-2001

SOURCE  
ORGANISM  
Corynebacterium glutamicum.  
Corynebacterium glutamicum  
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
Corynebacterium.  
REFERENCE  
AUTHORS Pompejus, M., Kroeger, B., Schroeder, H., Zelder, O., Haberhauer, G.,  
Kim, J. W., Lee, H. S. and Hwang, B. J.  
TITLE Corynebacterium glutamicum genes encoding metabolic pathway  
proteins  
JOURNAL Patent: WO 0166573-A 47 13-SEP-2001;  
BASF AKTIENGESellschaft (DE)  
FEATURES  
source  
1. 993  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
101..973  
/note="RXA01393"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC88506.1"  
/db\_xref="GI:15859120"  
translation="MNPQLDLSIIIDGSEFGASLALSISPSAVSRQVKALEHHVG  
RVLSRTQPAKATEAGVLAARQVLLQAEKTAQSLRLAEIPLTAINADLSLSTW  
FPFVNEVASWGGATLRLDEAHTLSLRGVDVLGAVTREANVAGCEVVELGTM  
HLAIATPSLRDAYMVDGLDWAAMPVLRFGKVDLQDRDLGRVDPGVRVSVIPS  
AEGFGEAIRRGLGWGLLPETQAPMLKAGEVILLDEIPIDFPMYQWRLLSRSLRL  
TDVVDDAAIEGLRP"  
BASE COUNT 233 a 256 c 268 g 236 t  
Query Match 41.8%; Score 993; DB 6; Length 993;  
Best Local Similarity 100.0%; Pred. No. 2.4e-233; Indels 0; Gaps 0;  
Matches 993; Conservative 0; Mismatches 0;  
Qy 1321 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 1380  
Db 1 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 60  
Qy 1381 AGTAAATGATGGTCTTAAACATGTTTAAATAGCTTCAATGAAACCCATTCAACTGGA 1440  
Db 61 AGTAAATGATGGTCTTAAACATGTTTAAATAGCTTCAATGAAACCCATTCAACTGGA 120  
Qy 1441 CACTTTGCTCTCAATCATATGATGAGGAGCTTGAAGCGGCTCTTTAGCCCTTTCCAT 1500  
Db 121 CACTTTGCTCTCAATCATATGATGAGGAGCTTGAAGCGGCTCTTTAGCCCTTTCCAT 180  
Qy 1501 TTCCCGCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACGTTGGGTGAGTGT 1560  
Db 181 TTCCCGCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACGTTGGGTGAGTGT 240  
Qy 1561 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGCTGAAGTCTTGTGCAAGCAGC 1620  
Db 241 GGTATCGCGCACCAACCGCGCAAGCAACCGAAGCGGCTGAAGTCTTGTGCAAGCAGC 300  
Qy 1621 CGGAAATGTTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGAGCGCTTGCTGA 1680  
Db 301 CGGAAATGTTGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGAGCGCTTGCTGA 360  
Qy 1681 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCACATGTTTCTCCCGTGT 1740  
Db 361 AATCCGTTAACCATCGCCATCAAGCAGATTCGCTATCCACATGTTTCTCCCGTGT 420  
Qy 1741 CAACGAGTAGCTTCTTTGGGTGGAGCAACGCTCAACGCTGCGCTTGAAGATGAAGCGCA 1800  
Db 421 CAACGAGTAGCTTCTTTGGGTGGAGCAACGCTCAACGCTGCGCTTGAAGATGAAGCGCA 480  
Qy 1801 CACATTATCTTGTGCGGCTGGAGATGTTTTAGAGCGGTAAACCGTGAAGCTAATCC 1860  
Db 481 CACATTATCTTGTGCGGCTGGAGATGTTTTAGAGCGGTAAACCGTGAAGCTAATCC 540  
Qy 1861 CGTGGCGGATGTCAGTAGTAGTGAAGCTTGAACCATGCGCACCTGCGCATTTGCCATTCGAAACCC 1920

```

Db 541 CGTGGGGGATGTGAAGTAGTAGAATTTGGAACCATCGCCACTTGGCCATTGCAACCCC 600
QY 1921 CTCAATTCGGGATGCCCTACATGTTGATGGAATACTAGATTGGCTCGGATGCCGCTCTT 1980
Db 601 CTCAATTCGGGATGCCCTACATGTTGATGGAATACTAGATTGGCTCGGATGCCGCTCTT 660
QY 1981 ACCTCTCGGTCCCAAGATGCTTCAAGACCGTAGCTGGAACCGGCGCTCGATGCTC 2040
Db 661 ACCTCTCGGTCCCAAGATGCTTCAAGACCGTAGCTGGAACCGGCGCTCGATGCTC 720
QY 2041 TGTGGGCGCAGGCGCTATTCATTTGTCCTCGGCGGAAGTTTGGTCAAGCAATTGC 2100
Db 721 TGTGGGCGCAGGCGCTATTCATTTGTCCTCGGCGGAAGTTTGGTCAAGCAATTGC 780
QY 2101 CCGAGGCGCTTGTGGGAGCTTCTTCCGAAACCAAGCTGCTCCCATGTCTAAAGCAGG 2160
Db 781 CCGAGGCGCTTGTGGGAGCTTCTTCCGAAACCAAGCTGCTCCCATGTCTAAAGCAGG 840
QY 2161 AGAAGTCACTCTCTCGATGAGATACCATTTGACACACCGCATTTATGCAAGCATGGCG 2220
Db 841 AGAAGTCACTCTCTCGATGAGATACCATTTGACACACCGCATTTATGCAAGCATGGCG 900
QY 2221 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATCGACCAATCGAGGG 2280
Db 901 CTTGGAATCTAGATCTCTAGCTAGACTCACAGACGCGCTGTTGATCGACCAATCGAGGG 960
QY 2281 ATTGGCGCTTAGTTACTTCTGAAAGGTTTCAG 2313
Db 961 ATTGGCGCTTAGTTACTTCTGAAAGGTTTCAG 993

RESULT 7
AX123540 AX123540 870 bp DNA linear PAT 11-MAY-2001
LOCUS Sequence 3456 from Patent EP1108790.
ACCESSION AX123540
VERSION AX123540.1 GI:14041028
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 870)
AUTHORS Nakagawa,S., Mizoguchi,H., Ando,S., Hayashi,M., Ochiai,K.,
Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.
TITLE Novel polynucleotides
JOURNAL Patent: EP 1108790-A 3456 20-JUN-2001;
KYOWA HAKKO KOGYO CO., LTD. (JP)
FEATURES
source Location/Qualifiers
1..870
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
BASE COUNT 192 a 236 c 245 g 197 t
ORIGIN

Query Match 36.6%; Score 870; DB 6; Length 870;
Best Local Similarity 100.0%; Pred. No. 4.4e-203;
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1421 ATGAACCCCAATCACTGGACACTTGTCTCAATCAATGATGAAGCAGCTTCGAAGGC 1480
Db 1 ATGAACCCCAATCACTGGACACTTGTCTCAATCAATGATGAAGCAGCTTCGAAGGC 60
QY 1481 GCCTCTTAGCCCTTCCATTTCCCTTCGGCGGTGAGTCAGCGCGTTAAAGCTCTCGAG 1540
Db 61 GCCTCTTAGCCCTTCCATTTCCCTTCGGCGGTGAGTCAGCGCGTTAAAGCTCTCGAG 120
QY 1541 CATCACTGGGTGAGTGTGGTATCGGCACCCCAACCGGCAAGCAACCGAAGCGGGT 1600
Db 121 GATCACTGGGTGAGTGTGGTATCGGCACCCCAACCGGCAAGCAACCGAAGCGGGT 180
QY 1601 GAAGTCTTGTGCAAGCAGCGCGGAAAATGTTGTCTGCAAGCAGAACTAAAGCGCAA 1660

```

```

Db 181 GAAGTCTTGTGCAAGCAGCGCGGAAAATGTTGTCTGCAAGCAGAACTAAAGCGCAA 240
QY 1661 CTATCTGGAGCGCTTGTCTGAAATCCCGTTAACCATCGCATCAACGACGAGATTGCTATCC 1720
Db 241 CTATCTGGAGCGCTTGTCTGAAATCCCGTTAACCATCGCATCAACGACGAGATTGCTATCC 300
QY 1721 ACATGTTTCTCTCCCGTGTTCACAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTG 1780
Db 301 ACATGTTTCTCTCCCGTGTTCACAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTG 360
QY 1781 CGCTTGGAGATGAGCGCACACATTTCTTCTGTCGCGCGTGGAGATGTTTTAGGAGCG 1840
Db 361 CGCTTGGAGATGAGCGCACACATTTCTTCTGTCGCGCGTGGAGATGTTTTAGGAGCG 420
QY 1841 GTAACCCGTGGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAACTTGGAAACCATGCGC 1900
Db 421 GTAACCCGTGGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAACTTGGAAACCATGCGC 480
QY 1901 CACTTGGCCATTGCAACCCCTCATTTGGCGGATGCCATCATGTTGATGGGAACTAGAT 1960
Db 481 CACTTGGCCATTGCAACCCCTCATTTGGCGGATGCCATCATGTTGATGGGAACTAGAT 540
QY 1961 TGGGCTGCGATGCGCGTCTTACGCTTCCGTCCTCAAGAGATGTCTTCAAGACCGTACCTG 2020
Db 541 TGGGCTGCGATGCGCGTCTTACGCTTCCGTCCTCAAGAGATGTCTTCAAGACCGTACCTG 600
QY 2021 GACGGCGCGTCTGATGCTCTGTCGGCGCGCAGCGCGTATTCATTGTCCTCGCGCGAA 2080
Db 601 GACGGCGCGTCTGATGCTCTGTCGGCGCGCAGCGCGTATTCATTGTCCTCGCGCGAA 660
QY 2081 GGTCTTGTGAGGCAATTCGCGAGCGCTTGTGGGGACTTCTTCCCGAAACCCCAAGCT 2140
Db 661 GGTCTTGTGAGGCAATTCGCGAGCGCTTGTGGGGACTTCTTCCCGAAACCCCAAGCT 720
QY 2141 GCTCCCATCTAAAGCAGGAGATGTCTTCTCGATGAGATACCCATTGACACACCG 2200
Db 721 GCTCCCATCTAAAGCAGGAGATGTCTTCTCGATGAGATACCCATTGACACACCG 780
QY 2201 ATGTATTGCAACGATGCGCGCTTGGAAATCTAGATCTCTAGTAGACTACAGACGCGTC 2260
Db 781 ATGTATTGCAACGATGCGCGCTTGGAAATCTAGATCTCTAGTAGACTACAGACGCGTC 840
QY 2261 GTTGTGAGCAATTCGAGGATTCGGGCT 2290
Db 841 GTTGTGAGCAATTCGAGGATTCGGGCT 870

RESULT 8
AX063771/c AX063771 822 bp DNA linear PAT 24-JAN-2001
LOCUS Sequence 53 from Patent WO0100843.
ACCESSION AX063771
VERSION AX063771.1 GI:12541483
KEYWORDS
SOURCE Corynebacterium glutamicum.
ORGANISM Corynebacterium glutamicum
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium.
REFERENCE 1 (bases 1 to 822)
AUTHORS Pompejus,M., Kroeger,B., Schroeder,H., Zelder,O. and Haberhauer,G.
TITLE Corynebacterium glutamicum genes encoding metabolic pathway
JOURNAL Patent: WO 0100843-A 53 04-JAN-2001;
BAS FANTINGSELSCHAFT (DE)
FEATURES
source Location/Qualifiers
1..822
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..802
/note="RXA01394"
/codon_start=1
CDS

```

/transl\_table=11  
/protein\_id="CAC25127.1"  
/db\_xref="GI:12541484"  
/translation="MEIPITGLLGLASLLSIPQNVLIKQIKREGILAVLVCLL  
SDVFLIAGTGLVDLLSNAPIVLDIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIIE  
ETPTVPDDTPLGGSAVADTDNRNRVREVSVDKORVWVKPMLMAIIVLTWLNPNAYLDA  
FVPIGVGAQYDGTGRWIFPAAGAFASLIWFLVPGFAAALSRPLSPKVRWVNVV  
AVMTALAIKLMNG"  
BASE COUNT 167 a 192 c 246 g 217 t  
ORIGIN

Query Match 34.6%; Score 822; DB 6; Length 822;  
Best Local Similarity 100.0%; Pred. No. 2.9e-191;  
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 629 GATTCCAAAACCCCGGAAACTTAACCATCAACATCAGTTTGTATGGCCATCGCGTCATC 688  
DB 822 GATTCCAAAACCCCGGAAACTTAACCATCAACATCAGTTTGTATGGCCATCGCGTCATC 763  
QY 689 ACAATGCCACGACGAGCTTGTATCAGCGGCACACCTTGGGGCTGGACAGCGGCGGTGAC 748  
DB 762 ACAATGCCACGACGAGCTTGTATCAGCGGCACACCTTGGGGCTGGACAGCGGCGGTGAC 703  
QY 749 AATGCTGCTGCGCGGAAACCCACAGCGGGAACAGATCAGGCTTGC CGGGAACCGGCCA 808  
DB 702 AATGCTGCTGCGCGGAAACCCACAGCGGGAACAGATCAGGCTTGC CGGGAACCGGCCA 643  
QY 809 GCGCGGAAATCCACCTCGGCTGTCGCGCTATTGCGCGGAGCGCGCGCGATTAACACA 868  
DB 642 GCGCGGAAATCCACCTCGGCTGTCGCGCTATTGCGCGGAGCGCGCGCGATTAACACA 593  
QY 869 AAGCGCTGCAAAATACGCAATTCGGGTTCACACAGCTCAGCAGCTTGGCATCAACATGGGC 928  
DB 582 AAGCGCTGCAAAATACGCAATTCGGGTTCACACAGCTCAGCAGCTTGGCATCAACATGGGC 523  
QY 929 TTATCCAAAACCCGCTGCTTATCGAGCTCACTTCAACCCGACCCGCGTTCGCGGTGCA 988  
DB 522 TTATCCAAAACCCGCTGCTTATCGAGCTCACTTCAACCCGACCCGCGTTCGCGGTGCA 463  
QY 989 GTGGCCACCGCGAACCAGCGGTGTCTATCGGGCAGCTTGGTCTGTCTTCTTCA 1048  
DB 462 GTGGCCACCGCGAACCAGCGGTGTCTATCGGGCAGCTTGGTCTGTCTTCTTCA 403  
QY 1049 ATGATCTGTGGCGCTTCCACTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTCA 1108  
DB 402 ATGATCTGTGGCGCTTCCACTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTCA 343  
QY 1109 CATTAACGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTC 1168  
DB 342 CATTAACGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTC 283  
QY 1169 GACAAAAGATCAACGCGCAAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 1228  
DB 282 GACAAAAGATCAACGCGCAAGGTGCGCGGATGAACAAAAGACGTCAGAAATTAACAC 223  
QY 1229 ACCGAGAAACCCGCAATGAGTCTTCTGCGCTTAATCTTCTTCTTCTTCTTCTTCTTCTTCA 1288  
DB 222 ACCGAGAAACCCGCAATGAGTCTTCTGCGCTTAATCTTCTTCTTCTTCTTCTTCTTCTTCA 163  
QY 1289 TGGGTCGATGACAGTAAAGACTGGCCCCCAAAAGCAGACCTGTATGAAGATTTC 1348  
DB 162 TGGGTCGATGACAGTAAAGACTGGCCCCCAAAAGCAGACCTGTATGAAGATTTC 103  
QY 1349 ATGATCACCATCTGACCTATGAAGTACTTAAGTAAATGATGGTCTTAAATCATGGTT 1408  
DB 102 ATGATCACCATCTGACCTATGAAGTACTTAAGTAAATGATGGTCTTAAATCATGGTT 43  
QY 1409 TAATATAGCTTCTTGAACCCCAATTCACCTGGACACTTTGCTC 1450  
DB 42 TAATATAGCTTCTTGAACCCCAATTCACCTGGACACTTTGCTC 1

RESULT 9

AX244059/c  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
CDS

CDS

1. 822  
/organism="Corynebacterium glutamicum"  
/db\_xref="taxon:1718"  
101..802  
/note="RXA01394"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAC88508.1"  
/db\_xref="GI:15859124"  
SDVFLIAGTGLVDLLSNAPIVLDIMRWGGIAYLLWFAVMAAKDAMTKVAPQIIIE  
ETPTVPDDTPLGGSAVADTDNRNRVREVSVDKORVWVKPMLMAIIVLTWLNPNAYLDA  
FVPIGVGAQYDGTGRWIFPAAGAFASLIWFLVPGFAAALSRPLSPKVRWVNVV  
AVMTALAIKLMNG"

BASE COUNT 167 a 192 c 246 g 217 t  
ORIGIN

Query Match 34.6%; Score 822; DB 6; Length 822;  
Best Local Similarity 100.0%; Pred. No. 2.9e-191;  
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCCAAAACCCCGGAAACTTAACCATCAACATCAGTTTGTATGGCCATCGCGTCATC 688  
DB 822 GATTCCAAAACCCCGGAAACTTAACCATCAACATCAGTTTGTATGGCCATCGCGTCATC 763  
QY 689 ACAATGCCACGACGAGCTTGTATCAGCGGCACACCTTGGGGCTGGACAGCGGCGGTGAC 748  
DB 762 ACAATGCCACGACGAGCTTGTATCAGCGGCACACCTTGGGGCTGGACAGCGGCGGTGAC 703  
QY 749 AATGCTGCTGCGCGGAAACCCACAGCGGGAACAGATCAGGCTTGC CGGGAACCGGCCA 808  
DB 702 AATGCTGCTGCGCGGAAACCCACAGCGGGAACAGATCAGGCTTGC CGGGAACCGGCCA 643  
QY 809 GCGCGGAAATCCACCTCGGCTGTCGCGCTATTGCGCGGAGCGCGCGCGATTAACACA 868  
DB 642 GCGCGGAAATCCACCTCGGCTGTCGCGCTATTGCGCGGAGCGCGCGCGATTAACACA 583  
QY 869 AAGCGCTGCAAAATACGCAATTCGGGTTCACACAGCTCAGCAGCTTGGCATCAACATGGGC 928  
DB 582 AAGCGCTGCAAAATACGCAATTCGGGTTCACACAGCTCAGCAGCTTGGCATCAACATGGGC 523  
QY 929 TTATCCAAAACCCGCTGCTTATCGAGCTCACTTCAACCCGACCCGCGTTCGCGGTGCA 988  
DB 522 TTATCCAAAACCCGCTGCTTATCGAGCTCACTTCAACCCGACCCGCGTTCGCGGTGCA 463  
QY 989 GTGGCCACCGCGAACCAGCGGTGTCTATCGGGCAGCTTGGTCTGTCTTCTTCTTCA 1048  
DB 462 GTGGCCACCGCGAACCAGCGGTGTCTATCGGGCAGCTTGGTCTGTCTTCTTCTTCA 403  
QY 1049 ATGATCTGTGGCGCTTCCACTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTCA 1108  
DB 402 ATGATCTGTGGCGCTTCCACTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTGTCTTCA 343  
QY 1109 CATTAACGGTAAGCGATGCGACCCAGCGCAATATATCGAGCAGATCGCGCGCGCATTC 1168



```
Db 342 CATACAGGTAAGGATGCGCCACCCAGCGCATATATCGAGCAGATCGCGCGGCGATTG 283
QY 1169 GACAAAAGATCAACGCCCAAGGTGCGCGCGATGAACAAAAGACGTCAGAAATTAACAC 1228
Db 282 GACAAAAGATCAACGCCCAAGGTGCGCGCGATGAACAAAAGACGTCAGAAATTAACAC 223
QY 1229 ACAGAGAAGACCGCAATGAGTCTTCGCGCTTAATTCCTTGTGTTAATCAGCAGTACATTC 1288
Db 222 ACAGAGAAGACCGCAATGAGTCTTCGCGCTTAATTCCTTGTGTTAATCAGCAGTACATTC 163
QY 1289 TCGCGTCCGATGACAGTAAAGACTGCGCCCAAGACGAGCTGTAATGAAGATTTC 1348
Db 162 TCGCGTCCGATGACAGTAAAGACTGCGCCCAAGACGAGCTGTAATGAAGATTTC 103
QY 1349 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATGTTGTTCTTAACATGGTT 1408
Db 102 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATGTTGTTCTTAACATGGTT 43
QY 1409 TAATATAGCTTCAAGACCCCATTCACACTGGACACTTTGCTC 1450
Db 42 TAATATAGCTTCAAGACCCCATTCACACTGGACACTTTGCTC 1

RESULT 10
AB083133/c
LOCUS
DEFINITION
Corynebacterium efficiens lysG, lysE genes for lysine export
transcriptional regulatory protein, Lysine exporter protein,
complete cds.
ACCESSION
AB083133
VERSION
AB083133.1 GI:20065738
KEYWORDS
Corynebacterium efficiens DNA.
SOURCE
Corynebacterium efficiens
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Corynebacteriaceae; Corynebacterium.
REFERENCE
1 Itaya, H., Kimura, E., Kawahara, Y. and Sugimoto, S.
lysG, lysE of Corynebacterium efficiens
Published Only in Database (2002)
2 (bases 1 to 1771)
Itaya, H., Kimura, E., Kawahara, Y. and Sugimoto, S.
Direct Submission
Submitted (04-APR-2002) Hiroshi Itaya, AJINOMOTO CO., INC,
Fermentation & Biotechnology Laboratories; 1-1, Suzuki-Cho,
Kawasaki-Ku, Kawasaki, Kanagawa 210-8681, Japan
(E-mail: hitoshi.itaya@ajinomoto.com, Tel: 81-44-244-7123 (ex. 4146),
Fax: 81-44-222-0129)
FEATURES
Source
1. 1771
/organism="Corynebacterium efficiens"
/db_xref="taxon:152794"
complement(47..919)
/gene="lysG"
complement(47..919)
/gene="lysG"
/codon_start=1
/transl_table=1
/product="Lysine export transcriptional regulatory
protein"
/protein_id="BAB88826.1"
/db_xref="GI:20065739"
/transl_table="NNPILHDTLLTI IDEGSPENASLALSISPSAVSORIKALEKSVG
RVLSVTPQAVAEAGVLTVOARKMALLOAETREOLAERLDEIPLTVAINADSLSTW
FPVPVAEVAHWGAVTLTLRVEDEAHTLSLLRSGVLGAVTREADPVAGCEVLRIGVMR
HLFVATPELARVTVDPQDVMVPLRPGNDVLDRLRLEGRVDGAVARRVSVPS
AESGFAVRILGVLGALLPEQAAPMLAAGDVQLDEKVDVDTFLYQWRKLESRLRLAR
TDAVDAARAGLRT"
981..1676
/gene="lysE"
981..1676
/gene="lysE"
```

```
/codon_start=1
/transl_table=1
/product="Lysine exporter protein"
/protein_id="BAB88827.1"
/db_xref="GI:20065740"
/transl_table="WRMEIFVTGLILGASILLAIAPONVLVIKQIKREGITAVIIV
CLLSDVLFITGLTGLVGLISDTAPILDLRWCGIAYLLWFAVWAARDALRARTBTF
VEHSEPVAAASAGSGVTTKORPLRITSGTQVWVWMLMAI VLTWLPNAYLLDAFV
FIGVGQYQYGETGRMIFAAAGAFASLVWFLVGVYGAALSRLSPRYWRWNIIGVAV
VLTGLAVKLIIMG"
BASE COUNT 294 a 603 c 545 g 329 t
ORIGIN
Query Match 33.6%; Score 798.8; DB 1; Length 1771;
Best Local Similarity 68.4%; Pred. No. 1.6e-185;
Matches 1158; Conservative 0; Mismatches 517; Indels 19; Gaps 3;
QY 650 TAACCCATCAACATCAGTTTGTATGCGCAATCGGTTCATCACAACTGCCACGACGAGTTG 709
Db 1675 TAACCCATCAGGATCAGTTTACGGCCAATCCGGTGAGCACCGGCCACACCTATGTTG 1616
QY 710 ATCCAGCGCCACACCTTGGGGCTGGACAGCGGGCGTGAACAATGCTGCTGCGCGCAACCC 769
Db 1615 ATCCAGCGCCACCGCGCGCTGGACAGGGGACGCGACAGTGGCGCGCGGTAACCG 1556
QY 770 ACAGCGGGAACAGATCAGGCTTGGCGGAACGCGGCCAGCGGGAANAATCACCCTCG 829
Db 1555 ACCAGTGGGAACAGACAGGCTGGCGGGAAGGACCGCGCAGCAAGATCCACCGACCG 1496
QY 830 GTGTGCGCGTATTCGGCGCGGCGGCGGATTAACACAAACGCGTCCAAATACGATTC 889
Db 1495 GTCTCCCGTACTGGGCTCCGACACCGCATGAAGACGAAGGCATCCAGGTAGGCATTG 1436
QY 890 GGTTTCAACAGCTCAGCAGCATGTCATCAATCGGGCTTTACCCAAACCCGCTGCTTA 949
Db 1435 GGATTCAGCAGGTCAGCAGCATGTCATGCGCATGAGTGGGCTGACCCAGACTTCCGGGTG 1376
QY 950 TCGACGCTCACCTCCACCGCACCGCGTTGCGGTGTCAGTGGCCACCGCCGAAACGCGCC 1009
Db 1375 CTTGATGTGATCGGAGCGCGGTGCTGTTTCTGTCGTCACGCCCGCGGAGGCGGAC 1316
QY 1010 AAGGCGTGTCTATCGGCGACGTTGTTCTGTTCTTCAATCATCTGTGCGGCTCCACC 1069
Db 1315 GCTGCGG----CAACGGTTTCGAATGCTC-----GACAAAGTTTACTCGGTG 1271
QY 1070 TTGTTTGTATGCGCGTCTTTCGTCGTCATGACGGCAAAACATAACAGGTAAAGCGTCCA 1129
Db 1270 CGGGCGCGCAGGCGCTGCGCGCGCCCATCATCCGCGAACCCACAGCAGGTAGGCGATCGC 1211
QY 1130 CCCAGCGCATTAATATCGAGCAGATCGCGCGGCGATTCGACAAAAGATCAACGCCCAAG 1189
Db 1210 CACAGCGCAGGATGTGAGAAATGATCGGGCGGTGTCGAGATCAGGCGCGACCCCGAGG 1151
QY 1190 GTGCGCGCATGAACAAAGACGTCAGAAATTAACACACAGAGAGACCCCAATGCT 1249
Db 1150 GTGCGCGGTTGAACAGGACCGTCCGACAGCAGACCATGATGATGACGCGCGGTGATG 1091
QY 1250 CTTTCGCGCTTAATTCCTTGTGTTTAAATCAGTACATCTCGCGGTCCGATGACAGTAA 1309
Db 1090 CCTCGCGTTGATGCGCTGTTTGTATCACCAGGACATCTGTGGGCGCGATGCCAACAGC 1031
QY 1310 AGACTGGCCCCCAAAAGCAGACCTGTAATGAAGATTTCCATGATACCATGATCGTGACCTAT 1369
Db 1030 AGACTGGCTCCCAACAAACAAACCGGTGAACAAAGATTTCCATGTCCCGCATGTTTCGCGC 971
QY 1370 GGNAGTACTTAAGTAAATGATGTTGTTCTTAACATGTTTAAATATAGCTTCATGACCCC 1429
Db 970 AAACCACTTAAGTAAATGATGAGACTGAACATGTTTAGAATTCCTTCATGAACCCG 911
QY 1430 ATTCAACTGGACACTTTGCTCTCAATCATTTGATGAAGGACGCTTCGAAAGCGCGCTCTTA 1489
Db 910 ATCCACCTGGACACCCCTCTGACCATCATCGAGGCGGCGAGCTTCGAGAACCGCTCCCTC 851
```

QY 1490 GCCCTTTCATTTCCCTCCGCGGTGAGTCAGCGCTTAAAGCTCTCGAGCATCACGTG 1549  
 Db 850 GCATGTGCGATCTCCCTCCGCGGTGAGTCAGCGCATCAAGGCCTCGGAAATCCGTC 791  
 QY 1550 GGTGCGAGTGTGTATCGCGCACCCCAACCGCCCAAGACACCGAGCGGTGAAGTCCIT 1609  
 Db 790 GGTGCGGTACTGTGTGCGCGCACCCCAACCGCGGTGTCACCGAGCGCGGAGGTGCTG 731  
 QY 1610 GTCAAGCAGCGCGGGAATGTTGCTCTCAAGCAGAACTAAAGCGCAACTATCTGGA 1669  
 Db 730 GTCCAGCGCGCCGAGATGCGCTGCTGAGCGCGAGACCGGTGAGCACTGCGGAA 671  
 QY 1670 CGCTTCTGAATCCCGTTAAACATCGCCATCAACGACATGCTGATCCACATGTTTT 1729  
 Db 670 CGCTTTGACGAGATCCCTCGACCGTGGCTATCAAGCCGATTCCTGTCCACTGCTTC 611  
 QY 1730 CTTCCCGTGTTCACGAGGTAGTCTTGGGTGGAGCAAGCTCAGCTGCGCTTGGA 1789  
 Db 610 CCGCGCTGTTCGCGAGGTGACATTTGGGTGCGGTCAACCTGACGCTGCGGTGGAG 551  
 QY 1790 GATGAAGCGCACATTTATCTTCTGCGCGCTGGAGATGTTTTAGAGCGGTAAACCGT 1849  
 Db 550 GACGAGCGCCACACCTCTGCTTGGCGCGGCTTCCGTGCTCGGGGCGGTGACCGCG 491  
 QY 1850 GAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAATTTGGAACCACTGCGCACTTGCGC 1909  
 Db 490 GAGCGGACCCCGTTGCGCGGTGTGAGTCTACGCTCGGGGTGATGAGACACCTGCGC 431  
 QY 1910 ATTGAACCCCTCATTTGCGGGATGCTACATGTTGATGGGAACCTAGATTGGGTGCG 1969  
 Db 430 GTGGCACCCCGGAATGAGGGCACGGGTACACCGTGGATGGCCAAACCCGACTGGGTGCGC 371  
 QY 1970 ATGCCGCTCTTACGCTTCGCTCCCAAGATGTCTTCAAGACCGTGAACCTGAGCGCGC 2029  
 Db 370 ATGCCGCTACTGCTGTTTCGCGCGGATGTGCTGAGGACCGCATCTCGAGGCGAG 311  
 QY 2030 GTCGATGCTCTGTGGGCGCAGCGCGTATCCATTGTCCGTCGCGGAGGTTTTGTT 2089  
 Db 310 GTCGAGGAGCGCTCGCAGCGCGCTGTCTCGGTGCTGCGGAGGTTTTGCGG 251  
 QY 2090 GAGGCAATTCGCGAGCGCTTGTGGGACCTTCTCCGAAACCCAGCTGCTCCGATG 2149  
 Db 250 GAGCGGTGCACTCGCGCTTGGGTGGGCGCTCTCGCGAGCGCCAGCGCGCGCGATG 191  
 QY 2150 CTAAAGCAGGAGAGTGTCTCTCGATGAGATACCCATTCACACCGCATGATGG 2209  
 Db 190 CTGGCTCGCGGATGTGTGCTGAGTGAAGGTGTGACACCCGCTGTATGG 131  
 QY 2210 CAACGATGCGCGCTGGAATCTAGATCTCTAGTAGACTCAAGACCGCTGTTGATGCA 2269  
 Db 130 CAACGATGCGCGCTGAGTCCAGGCTCTCGCCAGACTCAACACCGCTGTTGATGCG 71  
 QY 2270 GCATCGAGGATGCGGCTTAGTTACTTCTGAAAGTTGAGGTTTTCACTTCTTC 2329  
 Db 70 GCCCGCGGCTGTAG---AACCTAAACCTGGAACGTTCCGGTCTTGTACCTCTTC 15  
 QY 2330 GCCCGAGGAATTG 2343  
 Db 14 ACCGAGGATGTCG 1

RESULT 11  
 AX123539/c  
 LOCUS AX123539 708 bp DNA linear PAT 11-MAY-2001  
 DEFINITION Sequence 3455 from Patent EP1108790.  
 ACCESSION AX123539  
 VERSION AX123539.1 GI:14041027  
 KEYWORDS  
 SOURCE Corynebacterium glutamicum.  
 ORGANISM Corynebacterium glutamicum  
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 Actinomycetales; Corynebacterineae; Corynebacteriaceae;  
 Corynebacterium.

REFERENCE 1 (bases 1 to 708)  
 AUTHORS Nakagawa,S., Mizouchi,H., Ando,S., Hayashi,M., Ochiai,K.,  
 Yokoi,H., Tateishi,N., Senoh,A., Ikeda,M. and Ozaki,A.  
 TITLE Novel polynucleotides  
 JOURNAL Patent: EP 1108790-A 3455 20-JUN-2001;  
 KYOWA HAKKO KOGYO CO., LTD. (JP)  
 FEATURES  
 Location/Qualifiers  
 1..708  
 /organism="Corynebacterium glutamicum"  
 /db\_xref="taxon:1718" 180 t  
 BASE COUNT 134 a 173 c 221 g 180 t  
 ORIGIN  
 Query Match 29.8%; Score 708; DB 6; Length 708;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-163;  
 Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 652 ACCCATCAACATCAGTTTGATGCGCAATGCGTCTATCAAACTGCGACGACGTTGAT 711  
 Db 708 ACCCATCAACATCAGTTTGATGCGCAATGCGTCTATCAAACTGCGACGACGTTGAT 649  
 QY 712 CCAGCGCACACCTTGGGGCTGAGCAGCGGCGTGACATGCTGTGGCGCAAAACCCAC 771  
 Db 648 CCAGCGCACACCTTGGGGCTGAGCAGCGGCGTGACATGCTGTGGCGCAAAACCCAC 589  
 QY 772 CAGCGGGAACCAAGATCAGGCTTTCGCGCAACCGCGCAGCGGGAATAATCCACCGTCCGT 831  
 Db 588 CAGCGGGAACCAAGATCAGGCTTTCGCGCAACCGCGCAGCGGGAATAATCCACCGTCCGT 529  
 QY 832 GTCCGCGTATTGCGCGCGACCGCGCGGATAAACAAACGCGTCCAAATACGCAATTCGG 891  
 Db 528 GTCCGCGTATTGCGCGCGACCGCGCGGATAAACAAACGCGTCCAAATACGCAATTCGG 469  
 QY 892 GTTCAACAGGTGAGCAAGATTGCCATCAAATGGGCTTTACCCAAAACCGCTGCTTATC 951  
 Db 468 GTTCAACAGGTGAGCAAGATTGCCATCAAATGGGCTTTACCCAAAACCGCTGCTTATC 409  
 QY 952 GACGCTCACTCCACCGCACCGCGTTCGCGTGTCAGTGCCACCGCGCAACCGCCCAA 1011  
 Db 408 GACGCTCACTCCACCGCACCGCGTTCGCGTGTCAGTGCCACCGCGCAACCGCCCAA 349  
 QY 1012 AGCGGTGTCATCGGCGACGTTGTTCTGTTTCAATGATCTGTGGCGCTTCCACCTT 1071  
 Db 348 AGCGGTGTCATCGGCGACGTTGTTCTGTTTCAATGATCTGTGGCGCTTCCACCTT 289  
 QY 1072 GTTGTGTCAGGCTTTCGCTGCCATGACGCGAAACCAATACAGGTAGCGATGCCACC 1131  
 Db 288 GTTGTGTCAGGCTTTCGCTGCCATGACGCGAAACCAATACAGGTAGCGATGCCACC 229  
 QY 1132 CCAGCGCATATATCGACGACGATCGCGCGCATTTGGACAAAAGATCAACGCCCAAGGT 1191  
 Db 228 CCAGCGCATATATCGACGACGATCGCGCGCATTTGGACAAAAGATCAACGCCCAAGGT 169  
 QY 1192 GCCCGCGATGAACAAAAAGAGCTCAGAAATTAACACACGAGAAACCGCAATCAGTCC 1251  
 Db 168 GCCCGCGATGAACAAAAAGAGCTCAGAAATTAACACACGAGAAACCGCAATCAGTCC 109  
 QY 1252 TTCGCGCTTAATTCCTGTTTAATCAGCATACATTCCTGCGTCCGATGGACAGTAAAG 1311  
 Db 108 TTCGCGCTTAATTCCTGTTTAATCAGCATACATTCCTGCGTCCGATGGACAGTAAAG 49  
 QY 1312 ACTGCGCCCCAAAAGCAGACCTGTAATCAAGATTTCCATGATCACCAT 1359  
 Db 48 ACTGCGCCCCAAAAGCAGACCTGTAATCAAGATTTCCATGATCACCAT 1  
 RESULT 12  
 AX123538  
 LOCUS AX123538 627 bp DNA linear PAT 11-MAY-2001  
 DEFINITION Sequence 3454 from Patent EP1108790.  
 ACCESSION AX123538  
 VERSION AX123538.1 GI:14041026  
 KEYWORDS





AP003001/c  
LOCUS 345783 bp DNA linear BCT 15-MAY-2001  
DEFINITION Mesorhizobium loti DNA, complete genome, section 8/21.  
ACCESSION AP003001 BA000012  
VERSION AP003001.2 GI:14023393  
KEYWORDS  
SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.  
ORGANISM Mesorhizobium loti  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Phyllobacteriaceae; Mesorhizobium.  
REFERENCE  
1 (sites)  
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,  
Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,  
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,  
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimo, S., Sugimoto, M.,  
Takeuchi, C., Yamada, M. and Tabata, S.  
Complete genome structure of the nitrogen-fixing symbiotic  
bacterium Mesorhizobium loti  
DNA Res. 7 (6), 331-338 (2000)  
21082930  
2 (bases 1 to 345783)  
Kaneko, T.  
Direct Submission  
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research, Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/rhizobase/  
Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)  
On May 11, 2001 this sequence version replaced gi:11994976.  
COMMENT  
FEATURES  
source  
1. 345783  
/organism="Mesorhizobium loti"  
/strain="MAFF303099"  
/db\_xref="taxon:381"  
complement(6..884)  
/gene="mlr3011"  
complement(6..884)  
/gene="mlr3011"  
/codon\_start=1  
/transl\_table=11  
/product="transcription regulator"  
/protein\_id="BAB50000.1"  
/db\_xref="GI:14023394"  
/translation="MDRLMAMSTFIRVIETGSFSAARHLRGVSAVKAQLRERL  
GVRLSRTTGLSPDAGLSFVEGAKRAVDEADEAENAAAGTAAALTGILRTGTSTF  
GRHVPRFLKALDAPRLDVSAMDNDVLRGVDMVLKGSFPAHIGRKIA  
CRLVVAARSYLAREGEPIAPDELKSHAILFVRAGTSIRWTFRRGAEVSLAGGR  
RVTSAGTIREAVLSIGFTVGSSEWMPDAELRSGAVRPILTEWTLPPVELMALFPAGRK  
ASARATVFEFLAGSL"  
994..1740  
/gene="mlr3013"  
994..1740  
/gene="mlr3013"  
/codon\_start=1  
/transl\_table=11  
/product="short chain dehydrogenase or 3-ketoacyl-CoA  
reductase"  
/protein\_id="BAB50001.1"  
/db\_xref="GI:14023395"  
/translation="MGLNNKVAIVTGAARGIGAGIAKRLAAEGAAVNVVYATSR  
ERTVEIYAAGKATAVQGVAAAEVEMFEAAIAPGKPSILINNAATASFGTFEE  
ASEADYRRLFDVNLGTILTKAAHRHFPETGGSIVNIGTISSNPVPNTSISASKA  
AIDTLTSLRSRELGAIRNVIVAPGYTHTEMGVGTDFGNMLIAGVPLGQRGPK  
DIAPTVAFASDEAAWLTGERINASGNAR"  
1914..3371  
/gene="mlr3014"  
1914..3371  
/gene="mlr3014"  
/codon\_start=1  
/transl\_table=11  
/product="homoserimidine synthase"  
/protein\_id="BAB50002.1"  
/db\_xref="GI:14023400"  
/translation="MPKLDLRTGCRPVMSAYRAPAVTSAITRDTVTDVLLIINGISG  
/db\_xref="GI:14023396"  
/translation="MLDKMANENWVYGEITGPVVMIGFSGIGRGTLPILIERHFKFDK  
SRMTVLDPDRDTRKLLDGERGIAFVQEAETENNYKLLTPLLTNGGQGGQFCLNLSVDTG  
SVLMDRLCKGLVLIYDTPVEPWLFYFDKADANASRTNYALREAMIKEKHDPGGAT  
VSTCGANPGWVFWKQALVNLATDGLSEPAOEDREBGMKMKKAGKVGKIHAE  
RDTORTKPKMDVFWNTWVEGFISEGLQPAELGWTGHEKMMKPKNGKHKHGSKAAI  
YLEQGANTRVRSWCPTGPAQVGLLVTHNEAISADFFTVRSKKGKVOVPTCHVAYH  
PSNDAMLSDLDEMGAGKQPOVHHVLDENELVDGDELGLLVLYGHDKNAYWGSQSL  
REARKLAPYQATGCMQVTSVILAGMVALENPDAGIVEADEMDYKRCILDVQSPYLGVP  
KGYITWPLDRLRRPGLFEDLDLRSDFWQFRNLVR"  
3431..3931  
/gene="mlr3016"  
3431..3931  
/gene="mlr3016"  
/codon\_start=1  
/transl\_table=11  
/product="outer membrane lipoprotein"  
/protein\_id="BAB50003.1"  
/db\_xref="GI:14023397"  
/translation="MRPGAACVLRDARPFQHFGRSHRACQNRNRELDRCVWRGFOMTI  
ARKVLSIGMAGSFAMMLAECTSGPDAPPMAAPKGVGSGWIDAKGTSLSTFAGKFA  
TVADTQGLADGSYTMTGATSVETNGTSLIRQTPVPSPNCULLISTSQLNCTSSSQNF  
VLTRRT"  
4123..6243  
/gene="mlr3017"  
4123..6243  
/gene="mlr3017"  
/EC\_numbers="3.1.3.5"  
/codon\_start=1  
/transl\_table=11  
/product="5'-nucleotidase"  
/protein\_id="BAB50004.1"  
/db\_xref="GI:14023398"  
/translation="MKKLAATAALSVTLGLSAGASFPADYTLNLIHFNDRHSRIEGRN  
KYESTCSADEETKEGCEGAGRLITATAQERKLEQGNVLLNAGDSFGSLFTYTYK  
GTVEEFNLQMKPDAVTLGNHFDGGSALVYLDKAFPIVSANVMNDKSGAAGKI  
KPSIVVEVGQKIGIVGATVNDTPELASPGENIADIVESTAEVRKLKAGVNNK11  
AVTHIGNRERDIIAKIPGVVGVGSHSTLSNTDPAKAGPYPTWVNDPGCYKVPV  
QAASVSKYLGEFKVFDNDGVVKSAGSDPIFLDKSITPDPVAVLARIKELGAPTEALK  
KEVABTIDVIGSRENCCKEEMNLVDAILDRVKQGVEIIVISNGGGLRASIDKG  
TVTWGEVLTVLPFQNTATFQISGKDLVAGLESLQIEDGAGFPQVAGLKYSPDKS  
VAPNAGRVKSVEMENGAWTPIKPKDYLVTANNYVQGGGVKVFABRAKNAVDPG  
GLEQVADYLGARHPYTPKLDGRITETAAVAAAPAAEPAPKPAEATPAPAPAKPA  
EAAKPAEPKPAEAPAPKPAEAPKPAEAPKPAEAPKPAEAPKPAEAPKPAEAPKPA  
EAPAPAPAPVPAEPAPKPAEAPKPAEAPKPAEAPKPAEAPKPAEAPKPAEAPKPA  
IYEANKGQRHRLTPGATLTIPAK"  
6381..7442  
/gene="mlr3019"  
6381..7442  
/gene="mlr3019"  
/EC\_numbers="4.99.1.1"  
/codon\_start=1  
/transl\_table=11  
/product="ferrochelatase"  
/protein\_id="BAB50005.1"  
/db\_xref="GI:14023399"  
/translation="MTSAIVDPAKAPVGVTAGHAPAGAKGVLLVNLGTPTGTEF  
KPMRYLRFLSDPRVIELNKAIWYPIYGLVLTTRPKKGANTARINQRNESP  
TYTRAQSEKLAELGDLFDVTDWAMRYGNPSTASVAERLVAQCDRLSPFLYQV  
ITTTATANDOLFRAIMKLRAPAIRSVPPYAEPTVEALASSIERHLATLDFPEPV  
ITTSVHGIPKPYSDKGDVQAHLETTLRLKELGWDEBKLLITTFQSRFAGDEWLPQV  
DKTVEKLGKGVKSIAIVNPGFSVDCIETLDEIGREAAETPHAGGKNFAHICLND  
AEGMTVIAVMVRRELSGW"  
7615..8826  
/gene="mlr3020"  
7615..8826  
/gene="mlr3020"  
/note="hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB50006.1"  
/db\_xref="GI:14023400"  
/translation="MPKLDLRTGCRPVMSAYRAPAVTSAITRDTVTDVLLIINGISG





Search completed: April 27, 2003, 00:07:17  
Job time : 7473.94 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 20:02:03 ; Search time 3163.33 Seconds  
(without alignments)  
12154.292 Million cell updates/sec

Title: US-09-105-117K-3  
Perfect score: 2374  
Sequence: 1 agataactcttgggaagaa.....gtacacaccttcgaacatgg 2374

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST.\*

1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpi.\*  
7: em\_estro.\*  
8: em\_hct.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hct.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pln.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_man.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_other.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_rod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	84.6	3.6	657	14	BQ155035
C 2	73.2	3.1	3237	17	BH770958
C 3	72	3.0	860	17	AF075981
4	60.4	2.5	830	17	CNS01MQH
5	48.8	2.1	1101	17	CNS017SY
C 6	48.6	2.0	925	17	CNS0091P

C 7	44.4	1.9	644	14	BQ704105
8	44.4	1.9	664	17	CNS03K4J
9	42.8	1.8	532	10	AW927407
10	41.8	1.8	947	17	CNS077MV
11	41.8	1.8	1083	17	CNS079WN
12	41.6	1.8	558	14	BM712036
C 13	41.2	1.7	910	17	CNS006ON
14	41	1.7	617	14	BM729055
15	41	1.7	770	12	BE898495
16	41	1.7	861	12	BE740880
17	40.8	1.7	950	11	AY104298
C 18	40.6	1.7	436	17	AQ405681
C 19	40.2	1.7	613	10	BE498441
C 20	40	1.7	357	9	A1418751
C 21	40	1.7	431	9	AA452858
22	40	1.7	447	14	W91005
C 23	40	1.7	463	9	AA476500
C 24	40	1.7	504	9	A1159901
C 25	40	1.7	540	9	A1401605
C 26	40	1.7	564	9	A1820014
C 27	40	1.7	600	12	BG806166
C 28	40	1.7	916	17	CNS015QU
29	40	1.7	925	17	CNS0091P
C 30	39.8	1.7	1101	17	CNS00KK2
C 31	39.6	1.7	1667	11	AY108197
C 32	39.2	1.7	846	17	CNS010RJ
C 33	39.2	1.7	884	17	CNS006UJ
C 34	38.8	1.6	557	10	AW787854
C 35	38.8	1.6	1101	17	CNS017SY
C 36	38.6	1.6	436	10	BE025218
C 37	38.6	1.6	513	14	BQ818660
C 38	38.6	1.6	828	17	CNS026XT
39	38.4	1.6	650	10	BB576857
40	38.4	1.6	839	17	CNS004NB
C 41	38	1.6	212	9	AA150978
C 42	38	1.6	486	9	AA149286
C 43	38	1.6	903	13	BG975548
44	37.8	1.6	550	17	BH389688
45	37.8	1.6	804	17	CNS01HIJ

#### ALIGNMENTS

RESULT 1  
BQ155035/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BQ155035 657 bp mRNA linear EST 24-APR-2002  
NF075808IR1067 Irradiated Medicago truncatula cDNA clone  
NF075808IR 5', mRNA sequence.

BQ155035

BQ155035

BQ155035.1 GI:20292094

EST.

barrel medic.

Medicago truncatula

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;

Medicago.

1 (bases 1 to 657)

Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,

Flores,H.R., Inman,J.T., Weller,J.W. and May,G.D.

Expressed Sequence Tags from the Samuel Roberts Noble Foundation

Medicago truncatula irradiated library

Unpublished (2001)

Contact: May GD

Plant Biology Division

The Samuel Roberts Noble Foundation

2510 Sam Noble Parkway, Ardmore, OK 73402, USA

Tel: 580 221 7391

Fax: 580 221 7380

Email: gdmay@noble.org

Insert Length: 657 Std Error: 0.00

Plate: 075 row: E column: 08

Seq primer: TCACACAGGAACAGCTATGAC.

FEATURES  
source

Location/Qualifiers

1..657  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone\_lib="NF075E08IR"  
/clone\_lib="Irradiated"  
/tissue\_type="seedlings"  
/dev\_stage="seedling"  
/notes="Vector: Lambda Zap; Seedlings were exposed either to 100 Gy gamma or 0.5, 1, 5, or 10 kJ/m<sup>2</sup> UV irradiation. Gamma-irradiated samples were harvested at 6, 12, 24 and 48 hours after treatment. UV-irradiated samples were harvested 24 hours post-treatment. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRF' (Stratagene). Excised plasmids were plated using SOLR cells."

BASE COUNT 140 a 186 c 175 g 155 t 1 others

Query Match 3.6%; Score 84.6; DB 14; Length 657;  
Best Local Similarity 48.2%; Pred. No. 3.3e-13;  
Matches 237; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 1420 CATGAACCCCACTCACTGGACACTTTGCTCTCAATCATTTGATGAGGCGAGTTCGAAGG 1479  
DB 605 CCGGACTACAGACCATTAAGGCACTGGATGCGGTGATACGTGACGAGGATTTAGCG 546  
QY 1480 CGCTCTCTTACCCCTTCCATTTCCCTCGCGGTGAGTCAAGCGGTAAAGCTCTCGA 1539  
DB 545 CGCGGCAAAAGCTGTGATTAACAATCAGCGCTCTCACAGCGCATTAAGCAACTGGA 486  
QY 1540 GCATCAGTGGTGGAGTGTGATTCGCGACCAACCGCGGCAAGCAACCGAGCGGG 1599  
DB 485 AATATATGTTCCGGGAGCGCGCTGTGGTGGCTACCGGTACCGCGCGCGCGCAAGG 426  
QY 1600 TGAAGTCTTGTCAAGCAGCGGAAATGGTGTCTCAAGCAGAGAACTAAAGCGCA 1659  
DB 425 GCATAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366  
QY 1660 ACTATCGACGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1719  
DB 365 TGAACAAACGGTTCGACTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 306  
QY 1720 CATATGTTTCTCCGTTTCAAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCT 1779  
DB 305 GACGTGGTGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 246  
QY 1780 GCGTCTGGAAGATGAAGCGACACATTATCTTCTGCGCGGTGGAGATCTTTTAGGAGC 1839  
DB 245 GCAGGTAGAGATGAACCCGCACTCAGGAACGCTCTGCGCGCGGCGAGTGTGCGCGC 186  
QY 1840 GGTAAACCGGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGAATCTGGAACCAATGCG 1899  
DB 185 GGTGAGTATTCAACATCAGGCGCTGCGGAGTGTGCTTGTGATTAATAACTTGGTGGCTCGA 126  
QY 1900 CCACTTGGCCAT 1911  
DB 125 CTATCTGTTCGT 114

RESULT 2

BH770958/c

LOCUS BH770958 3237 bp DNA linear GSS 01-MAY-2002  
DEFINITION LMGtag686 MG1363 Random Sequence Tag Library Lactococcus lactis  
subsp. cremoris genomic, DNA sequence.

ACCESSION BH770958

VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BH770958.1 GI:20373915

GSS.  
Lactococcus lactis subsp. cremoris.  
Lactococcus lactis subsp. cremoris  
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
Lactococcus.

REFERENCE

1 (bases 1 to 3237)

AUTHORS

Bolotin,A., Ehrlich,S.D. and Sorokin,A.

TITLE

Studies of genomes of dairy bacteria Lactococcus lactis

JOURNAL

Sci. Aliments, (2002) In press

COMMENT

Contact: Sorokin A

Genetique Microbienne

INRA

CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

Tel: 33 1 34 65 25 16

Fax: 33 1 34 65 25 21

Email: sorokine@jouy.inra.fr

best homologue in strain IL1403 is nadR (98%)

Class: shotgun

High quality sequence start: 30

High quality sequence stop: 3207.

FEATURES

source

1..3237

/organism="Lactococcus lactis subsp. cremoris"

/strain="MG1363"

/db\_xref="taxon:1359"

/clone\_lib="MG1363 Random Sequence Tag Library"

/note="Vector: pSGM02; Site 1: SmaI; Library of

chromosomal fragments of L.lactis strain MG1363 was

prepared by partial AluI digestion or by sonication."

BASE COUNT 1058 a 577 c 613 g 980 t 9 others

ORIGIN

Query Match 3.1%; Score 73.2; DB 17; Length 3237;  
Best Local Similarity 49.8%; Pred. No. 1.3e-09;  
Matches 287; Conservative 0; Mismatches 263; Indels 26; Gaps 3;

QY 1 AGATACCTCTTTGGAGAAACCATGTACGATTGGTGCATCTTGGCTCTGGAAGGC 60  
DB 2511 AATACCCCTATTGAGAAACCATGGCGCTTTGAAACTGCTCTGGATAGTGGAAAGC 2452  
QY 61 TCTTTACGTGGGTATTCTTCTACGCTCCAGCTCACAGCGGCGGCTGAGTTTCAT 120  
DB 2451 ACTTTATGTGGGTGTCTTAATTATTTCTGCCAAGAAACCGAAGCTGCGCTTAGCGGC 2392  
QY 121 GCGGAGGAGGCTGCGCGCTTCTGATTCAATCAGCAAGCTATTCCATCATTAATCGTTG 180  
DB 2391 TGAAGAAGTTAGGCTTTAAACTCTTGATTTCATCAGCTCGTTATTCAATGTTAGTCGATG 2332  
QY 181 GGTGGAGGAACCGGGGATGACGGTGGAGAACTTGTTCAGTCACTGCGCAACATGGTCT 240  
DB 2331 GATTGAAGATGATTACAGAAACTTTGACAGAGGGG-----GGGAT 2289  
QY 241 TGGCGTCAATGCTTTCTACCACTTTCGCGAGGCTGCTCACGAGCAAAATATCTCGATGG 300  
DB 2288 AGAACAATGCTTTAGCGCTTTGTATCAAGACTCTTAAAGGAAATATTTGCAATGG 2229  
QY 301 AATTCAGAGGGTTC-----CCGCGCCAGCCAGGGTAAGTCCCTGTCTGAGGGCATGTT 354  
DB 2228 AATTCCTGAAATTCGGAATGCGTATCTCATTTATGCACTTTTGCAATGATGACAGTCT 2169  
QY 355 GAAGTGAACATATTGATATGCTCGCAAGCTCAATGACATCGCCCGAGACGGCGCA 414  
DB 2168 TACCACAGAAAGATTGGAACAAGCTTCAAGCCCTTAATGATTTAGCTCAAGTCGTGGACA 2109  
QY 415 GTCACTTGGCGAGATGGCGCTTTCATGGTGTGCGCGAGCAAGGAGAGTAGTCGCGCGCA 474  
DB 2108 ATCTCTAGCTCAATGGCATAAGCTTTGGGTTTACG---GGAAGAGCTGGCAAGTTCA 2052  
QY 475 TACCGTGAACAGTGAATGATTGGTGTCTGTCAGTTGAGCAGCTGGACAAACAGCCCTGA 534  
DB 2051 AGCTGTATACAGCTTTGATTGGGGCAAGTGGTCCGCAACAAATCATTTGAAAAATGTTGC 1992

```

QY 535 TTCACTCAACAACTTGGAGTTTCTGACGCCGAGTT 570
Db 1991 AGCTTTAGAACGTCGTGAATTTACTGACGAAGAATT 1956

RESULT 3
LOCUS AF075981
DEFINITION AF075981 Salmonella typhimurium LT2, Lambda DASH II linear GSS 29-AUG-2000
ACCESSION AF075981
VERSION AF075981.1 GI:3320851
KEYWORDS GSS.
SOURCE Salmonella typhimurium.
ORGANISM Salmonella typhimurium.
REFERENCE 1 (bases 1 to 860)
AUTHORS Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE Sample sequencing of a Salmonella typhimurium LT2 lambda library:
JOURNAL comparison to the Escherichia coli K12 genome
MEDLINE FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
COMMENT 92243757
Contact: McClelland M
Molecular Biology
Sidney Kimmel Cancer Center
3099 Science Park Road, San Diego, CA 92121, USA
Email: mclelland@lifesci.sdsu.edu
Class: shotgun.
FEATURES
source
1..860
/organism="Salmonella typhimurium"
/strain="LT2"
/db_xref="taxon:602"
/clone="390-T3"
/clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
/note="Vector: Lambda DASH II; sequenced using Li-Cor
sequencer"
BASE COUNT 212 a 199 c 242 g 207 t
ORIGIN

Query Match 3.0%; Score 72; DB 17; Length 860;
Best Local Similarity 54.7%; Pred. No. 1.8e-09;
Matches 220; Conservative 0; Mismatches 160; Indels 22; Gaps 3;

QY 52 TCGAAGGCTCTTTACGTGGTATTTCTTCTACGCTCCAGAGCTCACAGCGGAGCGGC 111
Db 18 TGGCAAGCGTTGTAGCTGGGATCTTAACATCTCTGCCGATCTGGCCAGACAAGCTAT 77

QY 112 TCAGTTTCATGGCGGAGGAGCGCTGCCGCTTCTTGATTCATCGACCAAGCTATTTCATCAT 171
Db 78 CGATATTCATGGAGGATCTCGGCAGCGCTTGCCTGATTCATCAGCCTAATATTCGCTTTT 137

QY 172 TAACTGTTGGGTGGAGGAACCGGGGATGACGGTGAGAACTTGTTCAGTCACTGAGTCCCAA 231
Db 138 TGAGCGTTGGGTAGAGGACGGGC-----TGCTGGCGCTGTTCAGGA 179

QY 232 CAATGCTCTTGGCGTCATGCTTCTCACCATTGCGCAGGCGCTGCTCACGGACAATA 291
Db 180 AAAAAAGCTCGGTAGTATTGCTTCTCGCGCGCTTGGGGCGGGCA-GCTCACTGACCGTAA 238

QY 292 TCTCGATGAAATCCAGAGGGTTCGCCG-----GCCAGCCAGGGTAACTGCTCTGTGAGG 348
Db 239 TTGGAATGTAATTCGGAGATTCGCGCGGGGAGCGGAGCCGTTTCTTAACCCAGA 298

QY 349 CATGTTGAACGTGAACATATTGATGTGTCGCAAGCTCAATGACATCGCCAGGAACG 408
Db 299 ACAGATTACCGCGCAAACTGGAATAAATGCTGCGCGTTGAAATGAATGAACTGGCTGCGGACG 358

QY 409 CGGGAGTCACTTGGCAGATCGCGCTTCATGGTGGCTGGG 450
Db 359 GGGTCAAAATTTGATCAATGGCGCTCGCTGGATGCTGGC 400

CNS01MQH 830 bp DNA linear GSS 14-JUN-2001
Anopheles gambiae GSS T7 end of clone 22E24 of NotreDamel library
from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
AL151258
AL151258.1 GI:7011737
GSS.
African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
1 (bases 1 to 830)
Genoscope.
Direct Submission
Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 830)
Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissensbach,J.
Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES
source
1..830
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="22E24"
/clone_lib="NotreDamel"
/note="end : T7"
BASE COUNT 207 a 231 c 227 g 156 t 9 others
ORIGIN

Query Match 2.5%; Score 60.4; DB 17; Length 830;
Best Local Similarity 50.7%; Pred. No. 4.2e-06;
Matches 142; Conservative 1; Mismatches 137; Indels 0; Gaps 0;

QY 1058 GCGGTTTCACCTTTTGTGTATGGCGTCTTTGGTCCATGACGGCAACCATACAGG 1117
Db 318 GCAGTTTCAGGTTACTGCTCATCGCCGTTTTTCAGCGCGCGAATCCGTACAGAGCAGA 377

QY 1118 TAAGCGATGCCACCCAGCGCATATATCGAGCACGATCGGGCGCATTCGACAAAGA 1177
Db 378 AACGCCAGCGCGCCCGAGTGACCAACCCAGCAGCCAGGAGACTGTCATCAGCAGCGG 437

QY 1178 TCAAGCCCAAGGTGCCGCGATGAACAAAGAGAGCTCAGAAATTAACACACAGAGAAGA 1237
Db 438 CTGCGCCCAAAATCCCGCACAGATCAGACCAATCACTTACCGCATAGAGGCA 497

QY 1238 ACCGAATGAGTCTTTCGGGCTTAATTCCTGTTTAAATCAACAGTACATTCGCGTCCG 1297
Db 498 ATCATCAGATGTAATGCGCGGATGCTGCTGATTCATCAGCAAGCATTTTTCGCGCGG 557

QY 1298 ATGGACATGAAGACTGCGCCCAAGAGCAGACTGTAA 1337
Db 558 AGGGAGAAATATAAATGCACCTAAGGCAAGCCCTTGAM 597

CNS017SY 1101 bp DNA linear GSS 26-JUL-1999
Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL108460

```





Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Zea.  
1 (bases 1 to 644)

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 946110 row: C column: 04.

## FEATURES

source

1. 644  
/organism="Zea mays"  
/cultivar="OH43"  
/db\_xref="taxon:4577"  
/clone\_lib="946 - tassels primordium prepared by Schmidt  
lab"  
/tissue type="tassels"  
/dev\_stage="just after the transition from vegetative to  
inflorescence development"  
/lab host="XLOLR"  
/note="Organ: tassels; Vector: HybriZAP; Site 1: ECGRI;  
Site 2: XhoI; George Chuck dissected immature tassels  
between 1mm and 3mm. Sharon Stanfield prepared the cDNA  
library in HybriZAP. Sample insert size range was 350 bp  
to 3 Kb with a 1 Kb average."  
118 a 239 c 180 g 107 t

BASE COUNT

ORIGIN

Query Match 1.9%; Score 44.4; DB 14; Length 644;  
Best Local Similarity 46.0%; Pred. No. 0.18;  
Matches 150; Conservative 0; Mismatches 176; Indels 0; Gaps 0;

QY 686 ATCAACATGTCACACGACGAGTGTATCCAGCGCACACCTTGGGTGACAGCGGGGT 745  
DB 39 ATCCCAATATCGCAACGCGATGTCTCGCTGCCCGCATGTGCTCTCGCGTGGCC 98  
QY 746 GACATGTCTGTCGCGCGAAGAACCCACAGCGGGAACAGATCAGGCTTCCCGCAACGCG 805  
DB 99 GCGCGCGCGCGCGAGCGGTCTACGCGGTCTAGCCAGCTCAGCGCGCGCGAAGCAG 158  
QY 806 CCAGCGCGGAAATCCACGCTCGGTGTGCGGTATGCGCGCGCGCGGATTAAC 865  
DB 159 CTACTCCCAATGACGCGCTCCGCGAGGGGCTCATCCCGGACTCCGCTCTTACAGC 218  
QY 866 ACAACGCGCTCCAAATACGATTCGGGTTCACACAGGTCTAGCAGGATTCGCAATCAACATG 925  
DB 219 TTCGAGGAGGACACGGCGGCTTGGATTCACCTCTCGCGGACCTGTACGTCCATTC 278  
QY 926 GCGTTTACCAACCGGTGTCTATCGAGCTCAGCTCAGCTCCACCGCACCGGTTGGCGGTG 985  
DB 279 GCGTCCCACTCGTCTACTACGAGAGGACCATACCGGCAAGCTCTCCAAAGGCGCCATC 338  
QY 986 TCAGTGGCCACCGCGAAGCGCCCA 1011  
DB 339 TCGGACCTCTCCGCGTCCAGGCCAA 364

RESULT 8

CN503K4J/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

CN503K4J 664 bp DNA linear GSS 17-MAY-2000  
Tetraodon nigroviridis genome survey sequence T7 end of clone  
032B06 of library G from Tetraodon nigroviridis, genomic survey  
sequence.  
AL247708  
AL247708.1 GI:7968720  
GSS: genome survey sequence.  
Tetraodon nigroviridis.

## ORGANISM

Tetraodon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
Tetraodontidae; Tetraodon.

## REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 664

/organism="Tetraodon nigroviridis"

/db\_xref="taxon:99883"

/clone="032B06"

/clone\_lib="G"

/note="Genoscope sequence ID : COBG032DA03LP1-end : T7"

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches 120;

Conservative

0;

Mismatches

129;

Indels

0;

Gaps

0;

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

RESULT 9

AW927407

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AW927407 532 bp mRNA linear EST 30-MAY-2000  
945001E08.Y1 945 - Mixed adult tissues from Walbot lab, same as 707  
(SK) Zea mays cDNA, mRNA sequence.

AW927407  
AW927407.1 GI:8102603  
EST.  
Zea mays.  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoidae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

1 (bases 1 to 532)  
Walbot, V.  
Maize ESTs from various cDNA libraries sequenced at Stanford  
University  
Unpublished (1999)  
Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Plate: 945001 row: E column: 08.

## FEATURES

## Source

1. .532

/organism="Zea mays"

/cultivar="W23"

/db\_xref="taxon:4577"

/clone\_lib="945 - Mixed adult tissues from Walbot lab,

same as 707 (SK)"

/issue type="tassel, kernal, silk, husk, root, leaf"

/dev stage="fully-grown"

/lab\_host="DH108"

/note="Organ: tassel, kernal, silk, husk, root, leaf;  
Vector: pGAD10; Site 1: EcoRI; cDNA library from fully  
differentiated maize tissues from an active Mutator plant.  
Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,  
root, leaf). Unidirectionally cloned. New library number  
given to library 707 for additional sequencing."

## BASE COUNT

## ORIGIN

Query Match 1.8%; Score 42.8; DB 10; Length 532;  
Best Local Similarity 45.7%; Pred. No. 0.5;  
Matches 149; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 686 ATCAAACTGCCACGACGAGCTTGATCCAGCGCCACACCTTGGGGCTGGACACGGCGGT 745

DB 7 ACCCCCAATATCCGACGGCAATGTCTCCGCTGCCCGCATCTGTCTCTCGCGCTCGCC 66

QY 746 GACAACTGCTCTCGCGGAAACCCACAGCGGACACAGATCAGGCTTCCGCGACGCG 805

DB 67 GCGCGCCCGCGGACGGGTGACGGCTGACGCCACGTCTCAGCGCGCGCGGACGAC 126

QY 806 CCAGCGCGGAAATCCACCCGCTCCGGTGTCTCGCGTATTTGCGCGCGGACGCGCGATAAAC 865

DB 127 CTACTCCCAAGTACGGCTCCCGAGGGGCTCATCCCGACTCCGTCGGCTCTACAGC 186

QY 866 ACAAACGGTCCAAATACGCAATTCGGGTTCACACGAGTCAAGCAGATTCGCAATCAACATG 925

DB 187 TTCAGCAGGCGCACGGCGGCTTCGAGATCCACTCGCGGACCTGTCTAGTCCACTTC 246

QY 926 GCGTTTACCAACCCGCTCTTATCGACGCTCACTCCACCGGACCGGTTGCGCGTG 985

DB 247 GGTCTCCACCTCTGTCTACTACGAGAGGACCAATACCGGCAAGCTCTCCAGGCGGCATC 306

QY 986 TCAGTGGCCACCGCGGACCGCCAA 1011

DB 307 TCGGACCTCTCCGGGCTCCAGGCCAA 332

## RESULT 10

## CNS077MW

## LOCUS

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

CNS077MW 947 bp DNA linear GSS 08-JUL-2001  
T3 end of clone BB0AA013All of library BB0AA from strain CBS 4732  
of Pichia angusta, genomic survey sequence.

AL432893

AL432893.1 GI:12216307

Pichia angusta.

Pichia angusta

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Pichia.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

## REFERENCE

```

RESULT 11
CNS079WN      1083 bp      DNA      linear      GSS 08-JUL-2001
LOCUS
DEFINITION
T3 end of clone XBB0AA001G01 of library XBB0AA from strain CBS 4732
of Pichia angusta, genomic survey sequence.
ACCESSION
AL435837
VERSION
AL435837.1  GI:12219250
KEYWORDS
GSS.
SOURCE
Pichia angusta.
ORGANISM
Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE
1 (bases 1 to 1083)
Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Liorente, B.,
Malpertuy, A., Neuvéglise, C., Ozier-Kalogeropoulos, O., Poirier, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE
20584711
PUBMED
11152876
REFERENCE
2 (bases 1 to 1083)
Blandin, G., Liorente, B., Malpertuy, A., Wincker, P., Artiguenave, P.
and Dujon, B.
Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta
FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE
20584723
PUBMED
11152888
REFERENCE
3 (bases 1 to 1083)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
secrefgenoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1..1083
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="XBB0AA001G01"
/clone_lib="XBB0AA"
/note="end : T3"
<40..>717
misc_feature
/note="similar to Saccharomyces cerevisiae ORF YPL088w [
similarity to aryl-alcohol dehydrogenases ]"
/evidence=not_experimental
BASE COUNT      326 a      214 c      234 g      308 t
ORIGIN
Query Match      1.8%; Score 41.8; DB 17; Length 1083;
Best Local Similarity 51.3%; Pred. NO. 1.3;
Matches 97; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
Qy 2 GATACCTCTTTCGAGAACACCATGACGATTCGGTACATTTGGCTCTCGAAGGCT 61
|||||
Db 163 GAGACACCATGTTGAGAACACGATGAGACGATTCGATGATTTGTTGAAGTCTGGAAGACT 222
|||||
Qy 62 CTTTACGTGGGTATTTCTTCTTACGGTCCAGAGCTCACACGGCGGCGGCTGAGTTCATG 121

```

```

Db 223 CGTTATTTGGGGCGCTCGACATGAGAGCTTATCAATTTATCGAATGCAACATGTTGCT 282
|||||
Qy 122 GCGGAGGAGGGCTGCCCGCTTCGTATTCATCAGCAAGCTATTCATCAATTAATCGTTGG 181
|||||
Db 283 GAAAGCATGGTTGGACAAAGTTTCATTTCCATCAAGCTATTTCCCTGTTCTTCAGA 342
|||||
Qy 182 GTGGAGGAA 190
|||||
Db 343 GAGGAAGAA 351
|||||
RESULT 12
LOCUS
BM712036      558 bp      mRNA      linear      EST 28-FEB-2002
DEFINITION
UI-E-DWI-ahc-g-16-0-UI-r1 UI-E-DWI Homo sapiens cDNA clone
UI-E-DWI-ahc-g-16-0-UI 5', mRNA sequence.
ACCESSION
BM712036
VERSION
BM712036.1  GI:19025294
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 558)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
9704477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
Clone Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com)
The following repetitive elements were found in this cDNA
sequence: 143-166, >GC-rich#low_complexity
Seq primer: M13 Reverse.
Location/Qualifiers
1..558
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-DWI-ahc-g-16-0-UI"
/clone_lib="UI-E-DWI"
/tissue_type="lens"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pTT3-Pac (Pharmacia) with a
modified polylinker; Site:1: EcoR I; Site:2: Not I;
UI-E-DWI is a normalized cDNA library containing the
following tissue(s): lens. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pTT3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CGATTACGGA. This library
was created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT      113 a      169 c      185 g      90 t      1 others
ORIGIN

```

Query Match 1.8%; Score 41.6; DB 14; Length 558;  
 Best Local Similarity 46.9%; Pred. No. 1.1; Mismatches 0; Gaps 0;  
 Matches 128; Conservative 0; Indels 145; Gaps 0;

QY 585 AGATTTCACACGCGCGCATCAATTTGGCGGAAGCCACCGATTTCACAAACCCGGG 644  
 DB 260 ACATCGCCACGAGGACTGATCAAGTGGCGCGGTCTCAACACGCTTCTACTGG 319  
 QY 645 AAACTAACCCATCAACATCAGTTTGTATGCGCCATGCGGTATCAACAATGCGCACGACGA 704  
 DB 320 AGAACCTGCACTTTCACATCGAGGCGAAGGACACGCACTACTTCTATCAAGACCAACGCG 379  
 QY 705 CGTTGATCAGCGCCACACTTGGGCTGACAGCGGGGTGACATGTCTGCGCCGA 764  
 DB 380 CCAGAGCGACCTGNGACGCTCGGTGTACACGCGCGCGCAAGGCGCTGGAGAACGCGCA 439  
 QY 765 AACCCACACGCGGACACCATCAGCTTCCGCGGAAGCGCGCGGCGGAAATCCACC 824  
 DB 440 TCAACGTGACGGTGTGCACTGACACCGTGTGTAACGCGGAGCGCGGAGTTGCGGG 499  
 QY 825 GTCGGGTGCGCGTATTGCGCGCGCAAGCGCCG 857  
 DB 500 ACGTGAGATGCACTTTCGCGCGCTGCGCGTGC 532

RESULT 13  
 CDS0060N/C  
 LOCUS  
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR14J21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
 ACCESSION AL065629  
 VERSION AL065629.1 GI:4944698  
 KEYWORDS GSS.  
 SOURCE Drosophila melanogaster.  
 ORGANISM Drosophila melanogaster.  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 1 (bases 1 to 910)  
 Direct Submission  
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoler in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
 source  
 1..910  
 /organism="Drosophila melanogaster"  
 /db\_xref="taxon:7227"  
 /clone="BACR14J21"  
 /clone\_lib="RPCI-98"  
 /note="end : T7"  
 BASE COUNT 202 a 63 c 112 g 198 t 335 others  
 ORIGIN  
 Query Match 1.7%; Score 41.2; DB 17; Length 910;  
 Best Local Similarity 14.8%; Pred. No. 1.8;  
 Matches 47; Conservative 150; Mismatches 119; Indels 2; Gaps 1;

QY 712 CCAGCGCCACACCTTGGGCTGGACAGCGGGCGTGCATCTGCTGCGCGCAACCCAC 771  
 DB 835 SVSSVSSASASASASASASVSSVSSVSSASRSRSASVSSAGSAVASAVAMACAS 776  
 QY 772 CAGCGGGAACAGATCAGGCTTGC CGCAACGCGCACGCGGAAATCCACCGTCCCGT 831  
 DB 775 WVASAVSMGASVVAARASAKSASMARVAAVARVASAAVASAAVAVASRVA 716  
 QY 832 GTCGCGTATTGCGCGCGACGCGCGCGGATTAACAACACGCGTCCAAATACGCAATCGG 891  
 DB 715 SASRVMASAGSSSSSSSSSSSSSSSSASMSVSVSAVSAASAAASARSASRA 656  
 QY 892 GTTCAACCA--GTCAGCAGCATGTCATCAATGCGCTTTACCCAAACCGCTGCTTA 949  
 DB 655 SAVMAVAASASAVASGSSSVSCSSRSRSCGSSSSSSSSSSSSSSSSSSSSSSSS 596  
 QY 950 TCGACGCTCACTCCACCGCGCGGTTGCGCGTGCAGTGGCCACCGCGCAACCGCC 1009  
 DB 595 ASCASSSSRSRSCGCGSVGCGSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 536  
 QY 1010 AAAGCGGTGTATCGCGG 1027  
 DB 535 CCRSCSCGCGCCSSCC 518

RESULT 14  
 BM729055  
 LOCUS  
 DEFINITION UI-E-E01-aiw-a-06-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone  
 UI-E-E01-aiw-a-06-0-UI 5', mRNA sequence.  
 ACCESSION BM729055  
 VERSION BM729055.1 GI:19050388  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 617)  
 REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 AUTHORS Normalization and subtraction: two approaches to facilitate gene  
 TITLE discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704477  
 COMMENT Contact: Soares, MB  
 Program for Rat Gene Discovery and Mapping  
 University of Iowa  
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: msoares@blue.weeg.uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 148-171, >GC rich#Low\_complexity  
 Seq primer: M13 Reverse

FEATURES  
 source  
 1..617  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-E01-aiw-a-06-0-UI"  
 /clone\_lib="UI-E-E01"  
 /tissue\_type="fetal eye"  
 /dev\_stage="fetal"  
 /lab\_host="DHI0B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-E01 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was

constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7733-fac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

```

BASE COUNT      124 a  189 c  211 g  93 t
ORIGIN
Query Match      1.7%; Score 41; DB 14; Length 617;
Best Local Similarity 46.9%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 585 AGATTTCCACGACGCGGCATCAACATTTGGGCGAAGGCCACCGATTCAAAACCCGCG 644
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 ACATCGCAACGAGGACTGCATCAAGGTGGCGCGCTGCTCAACAGCCCTTCTACCTGG 324
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 AAACTAACCCATCAACATCAGTTTGATGCGCAATGGGTTCATCAACATGCGCACGACGA 704
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 AGAACCTGCACCTTCCACATCGAGCGCAAGGACACGCACTTCTATCAAGACCACCAACGC 384
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 705 CGTTGATCCAGCGCCACACCTTGGGGCTGGACAGCGCGGTGCAATGCTGCTGCGCGGA 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 CCAGAGCGACTTGGGACACGCTGCGGTTGACAGCGCGCCCAAGGCGCTGGAGAACGGCA 444
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 765 AACCCACGACGCGGAACAGATCAGGTTGCGCGCAACGCGCGCGGCAAAATCCACC 824
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 445 TCAACGTGACGGTGTGCACTGCACACAGGTGTGTAAGCGCAGGACGCGAGGTTCGCGG 504
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 825 GTCCGTGTGCGCGTATGTCGCGCGCGCGCGCG 857
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 ACGTGGAGATGCAGTTTCGGCGCGCTGCGCGTGC 537
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 15
BE898495      770 bp  mRNA  linear  EST 29-SEP-2000
LOCUS        601681396F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3951622 5',
DEFINITION   mRNA sequence.
ACCESSION    BE898495
VERSION      BE898495.1 GI:10365029
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 770)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: DCTB/DTF
              cDNA Library Preparation: Ling Hong/Rubin Laboratory
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Incyte Genomics, Inc.
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
              Plate: LLCW819 row: p column: 23
              High quality sequence stop: 765.
              Location/Qualifiers
                1..770
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:3951622"
                  /clone_lib="NIH_MGC_9"
                  /tissue_type="adenocarcinoma cell line"

```

```

FEATURES
source

```

```

/lab host="DH10B (phage-resistant)"
/note="organ: ovary; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      147 a  245 c  277 g  101 t
ORIGIN
Query Match      1.7%; Score 41; DB 12; Length 770;
Best Local Similarity 46.9%; Pred. No. 1.9;
Matches 128; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```

```

QY 585 AGATTTCCACGACGCGGCATCAACATTTGGGCGAAGGCCACCGATTCAAAACCCGCG 644
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 113 ACATCGCCAAACGAGGACTGCATCAAGGTGGCGCGCTGCTCAACAAAGCCTTCTACCTGG 172
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 645 AAACTAACCCATCAACATCAGTTTGATGCGCAATGGGTTCATCAACATGCGCACGACGA 704
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 AGAACCTGCACCTTCCACATCGAGCGCAAGGACACGCACTTCTATCAAGACCACCAACGC 232
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 705 CGTTGATCCAGCGCCACACCTTGGGGCTGGACAGCGCGGTGCAATGCTGCTGCGCGGA 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 CCAGAGCGACTTGGGACACGCTGCGGTTGACAGCGCGCGCAAGGCGCTGGAGAACGGCA 292
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 765 AACCCACGACGCGGAACAGATCAGGTTGCGCGCAACGCGCGCGGCAAAATCCACC 824
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 TCAACGTGACGGTGTGCGAGTCCACACGGTGTGTAAGCGCAGGACGCGAGGTTCGCGG 352
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 825 GTCCGTGTGCGCGTATGTCGCGCGCGCGCGCG 857
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 ACGTGGAGATGCAGTTTCGGCGCGCTGCGCGCTGC 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: April 27, 2003, 00:51:42  
Job time : 3177.33 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2003, 17:06:23 ; Search time 522.424 Seconds  
(without alignments)  
10233.535 Million cell updates/sec

Title: US-09-105-117K-3  
Perfect score: 2374  
Sequence: 1 agatactctttggaagaaa.....gtaacaccttcagaaatgg 2374

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	2374	100.0	2374	18 AAT96816	DNA encoding LysG,
2	2374	100.0	349980	22 AAF68528	C glutamicum codin
3	993	41.8	993	22 AAF71777	Corynebacterium gl
4	993	41.8	993	23 AAS96096	C. glutamicum gene
5	870	36.6	870	22 AAF68421	C glutamicum codin
6	822	34.6	822	22 AAF71779	Corynebacterium gl
7	822	34.6	822	23 AAS96098	C. glutamicum gene
8	708	29.8	708	22 AAF68420-	C glutamicum codin
9	698.6	29.4	1568	22 AAF45375	C. thermoaminogene

10	627	26.4	627	22	AAH68419	C glutamicum codin
11	485.2	20.4	993	22	AAF68077	Corynebacterium gl
12	121.2	5.2	15239	17	AAT33536	BCG deletion regio
C 13	121.2	5.1	4403765	22	AAI99683	Mycobacterium tube
C 14	121.2	5.1	4411529	22	AAI99682	Mycobacterium tube
C 15	101	4.3	1200	22	AAF71729	Corynebacterium gl
C 16	100.8	4.2	5541	15	AAQ55755	Escherichia coli g
17	93.2	3.9	988	24	ABK72786	Bacillus lichenifo
18	82.6	3.5	894	23	AAS52558	E. coli DNA for ce
19	79.4	3.3	23128	23	AAS59552	Propionibacterium
C 20	78	3.3	1077	22	AAH66360	C glutamicum codin
21	76.2	3.2	894	23	AAS56007	Salmonella typhi D
22	73	3.1	903	23	AAS51548	Pseudomonas aerugi
23	68.8	2.9	2365589	24	ABA90521	Genomic sequence o
24	64.4	2.7	1452	23	AAS88994	DNA encoding novel
25	64	2.7	1277	24	AAD31869	Lactobacillus rham
C 26	60.4	2.5	636	21	AAAS2691	Escherichia coli y
27	57.6	2.4	762	23	AAS54006	Klebsiella pneumon
C 28	47.4	2.0	7521	23	AAS71378	DNA encoding novel
C 29	47.4	2.0	7521	23	AAS94251	DNA encoding novel
C 30	45.4	1.9	45613	22	AAF28535	Genomic fragment #
C 31	45.2	1.9	4915	20	AAH84317	Human breast cell
32	43.8	1.8	1973	22	ABA44805	Stealth virus nucl
33	43.8	1.8	1973	22	ABA55261	Human foetal liver
34	43.8	1.8	1973	22	ABA25005	Probe #3471 for ge
35	43.8	1.8	1973	22	AAK03514	Human brain expres
36	43.8	1.8	1973	22	AAK28970	Human bone marrow
37	43.8	1.8	1973	22	AAI13556	Probe #3489 for ge
38	43.8	1.8	1973	22	AAI34918	Probe #3604 used t
39	43.8	1.8	1973	22	AAI03446	Probe #3437 used t
40	43.8	1.8	1973	24	ABS03504	Human genome-deriv
C 41	41.6	1.8	2842	21	AAD00334	Rice raffinose syn
42	41	1.7	2157	22	AAH14096	Human cDNA sequenc
43	41	1.7	3270	22	AAH14671	Human cDNA sequenc
44	41	1.7	3614	22	AAH14183	Human cDNA sequenc
C 45	40.4	1.7	1075	22	ABA48937	Human breast cell

## ALIGNMENTS

RESULT 1  
AAT96816/c  
ID AAT96816 standard; DNA; 2374 BP.  
XX  
AC AAT96816;  
XX  
DT 12-MAR-1998 (first entry)  
DE DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.  
XX  
KW LysG, LysE; ORF3; lysine transport; regulatory protein; export protein;  
KW Microbial production; amino acid; animal feed additive; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
FH Key Location/Qualifiers  
FT CDS complement (82..954)  
FT FT /\*tag= a  
FT FT /label= LysG  
FT FT 1016..1726  
FT FT /\*tag= b  
FT FT /label= LysE  
FT FT complement (1723..2373)  
FT FT /\*tag= c  
FT FT /label= orf3

DE19548222-A1.

26-JUN-1997.

22-DEC-1995; 95DE-1048222.



PR 22-DEC-1995; 95DE-1048222.  
 XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.  
 XX  
 XX Eggeling L, Sahm H, Vrljic M;  
 XX  
 XX WPI; 1997-333867/31.  
 DR P-PSDB; AAW37714-16.  
 DR  
 XX  
 PT Increasing microbial production of amino acids, especially lysine -  
 PT by improving export carrier activity or corresponding gene  
 PT expression, also new export and regulatory genes from  
 PT Corynebacterium  
 XX  
 PS Claim 23 and 26; Page -: 16pp; German.  
 XX  
 CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,  
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory  
 CC protein and an export protein, respectively. Microbial production of  
 CC amino acids (A) is improved by increasing the export-carrier activity  
 CC and/or the export gene expression in a microorganism that produces (A).  
 CC The method is specifically used to increase production of lysine,  
 CC used as an animal feed additive. Other (A) are variously useful as  
 CC pharmaceuticals, condiments and intermediates for fine chemicals.  
 CC This method increases the amount of (A) secreted into the culture medium.  
 CC Export of (A) has been found to depend on a single gene.  
 CC NB. This sequence has been created from the information given in table 2  
 CC of the specification.  
 XX  
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;  
 Query Match 100.0%; Score 2374; DB 18; Length 2374;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGATACCTCTTGAAGAACCATGTACGATTCGCGTACATGTTCCGCTCGAAGGC 60  
 DB 2374 AGATACCTCTTGAAGAACCATGTACGATTCGCGTACATGTTCCGCTCGAAGGC 2315  
 QY 61 TCTTTACGTGGGTATTTCTTCTACCGTCCAGAGCTCACAGCGAGCGCGCTGAGTTTCA 120  
 DB 2314 TCTTTACGTGGGTATTTCTTCTACCGTCCAGAGCTCACAGCGAGCGCGCTGAGTTTCA 2255  
 QY 121 GCGGAGAGGGCTGCGCGCTTCTGATTCATCAGCAGCTATTCATCATTAATCGTTG 180  
 DB 2254 GCGGAGAGGGCTGCGCGCTTCTGATTCATCAGCAGCTATTCATCATTAATCGTTG 2195  
 QY 181 GGTGAGGAACCGGGCGATGACGCTGAGAACTTGTTCAGTTCAGTCCGCAACATGTCT 240  
 DB 2194 GGTGAGGAACCGGGCGATGACGCTGAGAACTTGTTCAGTTCAGTCCGCAACATGTCT 2135  
 QY 241 TGGCGTCATTGCTTTCTCACCACCTTGGCGAGGGCTGCTCACGAGCAATATCTCGATGG 300  
 DB 2134 TGGCGTCATTGCTTTCTCACCACCTTGGCGAGGGCTGCTCACGAGCAATATCTCGATGG 2075  
 QY 301 AATTCCAGAGGGTTCGCGCGCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACGT 360  
 DB 2074 AATTCCAGAGGGTTCGCGCGCAGCAGGTAAGTCCCTGTCTGAGGCGATGTTGAACGT 2015  
 QY 361 GAACAATATTGATGTTGTCGCAAGCTCAATGATCGCCAGCAACCGGGAGTCAC 420  
 DB 2014 GAACAATATTGATGTTGTCGCAAGCTCAATGATCGCCAGCAACCGGGAGTCAC 1955  
 QY 421 TGGCGAGATGGCGCTTGATGGTGTGCTGCGCGAAGAGGAGTACGCGCGGATACCGT 480  
 DB 1954 TGGCGAGATGGCGCTTGATGGTGTGCTGCGCGAAGAGGAGTACGCGCGGATACCGT 1895  
 QY 481 GACAGTGCATTGATGGTGTGCTGCTGATGAGCAGTGAACAGCCTTGATTCAT 540  
 DB 1894 GACAGTGCATTGATGGTGTGCTGCTGATGAGCAGTGAACAGCCTTGATTCAT 1835  
 QY 541 CAACAACTTGAGTGTTCGTACGCGCGAGTTGGAGGCGATCGATCAGATTTCCACGACGC 600

DB 1834 CAACAACTTGAGTGTTCGTACGCGCGAGTTGGAGGCGATCGATGAGATTTCCACGACGC 1775  
 QY 601 CGGCATCAACATTTGGGCGAAGGCCACCGATTCCAAAACCCGCGAAAACCTAAACCCATCAA 660  
 DB 1774 CGGCATCAACATTTGGGCGAAGGCCACCGATTCCAAAACCCGCGAAAACCTAAACCCATCAA 1715  
 QY 661 CATCAGTTTGTATGGCCCAATCGGTTCATCACTCACTCACTCACTCACTCACTCACTCACTCA 720  
 DB 1714 CATCAGTTTGTATGGCCCAATCGGTTCATCACTCACTCACTCACTCACTCACTCACTCACTCA 1655  
 QY 721 CACCTTTGGGGCTGAGACAGCGGGCGTGAACAATCTGTGCGCGGAAAACCCACAGCGGAAA 780  
 DB 1654 CACCTTTGGGGCTGAGACAGCGGGCGTGAACAATCTGTGCGCGGAAAACCCACAGCGGAAA 1595  
 QY 781 CCAGATCAGGCTTGGCGGAAACGCGCCAGCGGCGGAAAATCCACCGTCCGGTGTCCCGTA 840  
 DB 1594 CCAGATCAGGCTTGGCGGAAACGCGCCAGCGGCGGAAAATCCACCGTCCGGTGTCCCGTA 1535  
 QY 841 TTGGCGCGCGACCGCGCGATTAACACAAACCGCTCCAAATACGATTCGGGTTCACCA 900  
 DB 1534 TTGGCGCGCGACCGCGCGCGATTAACACAAACCGCTCCAAATACGATTCGGGTTCACCA 1475  
 QY 901 GGTACAGCAGATTGCCATCAACATGCGCTTTACCCAAACCCGCTGCTTATCGACGCTCAC 960  
 DB 1474 GGTACAGCAGATTGCCATCAACATGCGCTTTACCCAAACCCGCTGCTTATCGACGCTCAC 1415  
 QY 961 CTCACCCCGACCCCGTTCGCGGTGTGAGTGGCCACCGCGGAAACCCCAAGGGGTGTC 1020  
 DB 1414 CTCACCCCGACCCCGTTCGCGGTGTGAGTGGCCACCGCGGAAACCCCAAGGGGTGTC 1355  
 QY 1021 ATCGGCGACGGTGTGTTCTGTTCTTCATGATCTGTGGCGCTTCACCTGTTGTCAT 1080  
 DB 1354 ATCGGCGACGGTGTGTTCTGTTCTTCATGATCTGTGGCGCTTCACCTGTTGTCAT 1295  
 QY 1081 GGGCTCTTTTCGCTGATCAACGCGAAACCATAAACAGGTAAAGCGATGCCACCCAGCGCAT 1140  
 DB 1294 GGGCTCTTTTCGCTGATCAACGCGAAACCATAAACAGGTAAAGCGATGCCACCCAGCGCAT 1235  
 QY 1141 AATATCGACACGATCGCGCGCATTTGCAAAAAGATCAACGCCCAAGGTGCGCGCAT 1200  
 DB 1234 AATATCGACACGATCGCGCGCATTTGCAAAAAGATCAACGCCCAAGGTGCGCGCAT 1175  
 QY 1201 GAACAAAAGAGCTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTTCGGCTT 1260  
 DB 1174 GAACAAAAGAGCTCAGAAATTAACACACGAGAGAACCGCAATGAGTCTTTCGGCTT 1115  
 QY 1261 AATTCTTGTGTTAATCACCAGTACATTTCTGCGGTCCGATGAGCAAGTAAAGACTTGGCCGC 1320  
 DB 1114 AATTCTTGTGTTAATCACCAGTACATTTCTGCGGTCCGATGAGCAAGTAAAGACTTGGCCGC 1055  
 QY 1321 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 1380  
 DB 1054 CAAAAGCAGACCTGTAAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 995  
 QY 1381 AGTAAATGATGTTCTTAAACATGTTTAAATATAGTTCATGACCCCAATCAATGGA 1440  
 DB 994 AGTAAATGATGTTCTTAAACATGTTTAAATATAGTTCATGACCCCAATCAATGGA 935  
 QY 1441 CACTTGTCTCTCAATGATGAGGAGCTTTCGAAAGCGCTCTTAGCCCTTTCCAT 1500  
 DB 934 CACTTGTCTCTCAATGATGAGGAGCTTTCGAAAGCGCTCTTAGCCCTTTCCAT 875  
 QY 1501 TTCCCTCTCGCGGTGAGTCAGCGGTTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 1560  
 DB 874 TTCCCTCTCGCGGTGAGTCAGCGGTTTAAAGCTCTCGAGCATCACGTGGGTGAGTGT 815  
 QY 1561 GGTATCGCGACCCCAACCGGCCCAACCGAGCGGTGAGTCTTGTGCAAGCAGC 1620  
 DB 814 GGTATCGCGACCCCAACCGGCCCAACCGAGCGGTGAGTCTTGTGCAAGCAGC 755  
 QY 1621 GCGGAAAATGGTGTCTGCTGCAAGCAGAAAATAAAGCGCAACTATCTGGACGCCCTTCTGA 1680  
 DB 754 GCGGAAAATGGTGTCTGCTGCAAGCAGAAAATAAAGCGCAACTATCTGGACGCCCTTCTGA 695

QY 1681 AATCCCGTTAAACCATCGCCATCAACGAGATTCGCTATCAATGTTTCTCCCGTGT 1740  
DB |||||  
QY 694 AATCCCGTTAAACCATCGCCATCAACGAGATTCGCTATCAATGTTTCTCCCGTGT 635  
DB |||||  
QY 1741 CAACGAGTAGCTCTTGGGGTGAGCAAGCTCAGCTGCGCTTGGAGATGAAGCGCA 1800  
DB |||||  
QY 634 CAACGAGTAGCTCTTGGGGTGAGCAAGCTCAGCTGCGCTTGGAGATGAAGCGCA 575  
DB |||||  
QY 1801 CACATTATCTTCTGCTCGCGGTGAGATGTTTGGAGCGGTAAACCGTGAGCTAATCC 1860  
DB |||||  
QY 574 CACATTATCTTCTGCTCGCGGTGAGATGTTTGGAGCGGTAAACCGTGAGCTAATCC 515  
DB |||||  
QY 1861 CBTGGCGGATGTGAATAGTAGAATCTTGGAAACCATCGCCACTTGGCCATTGCAACCCCG 1920  
DB |||||  
QY 514 CBTGGCGGATGTGAATAGTAGAATCTTGGAAACCATCGCCACTTGGCCATTGCAACCCCG 455  
DB |||||  
QY 1921 CTCATTGCGGATGCTTACATGTTTCATGGGAACTAGATTGGCTGCGATGCCCTCTT 1980  
DB |||||  
QY 454 CTCATTGCGGATGCTTACATGTTTCATGGGAACTAGATTGGCTGCGATGCCCTCTT 395  
DB |||||  
QY 1981 AGCTTTCGGTCCCAAGATGTGCTTCAAGACCGTGAACGCGCGCTGATGCTCC 2040  
DB |||||  
QY 394 AGCTTTCGGTCCCAAGATGTGCTTCAAGACCGTGAACGCGCGCTGATGCTCC 335  
DB |||||  
QY 2041 TGTGGGCGGAGCGGCTATCCATTGTCCTGCGGGAAGTGTGTTGGTGAGGCAATTGC 2100  
DB |||||  
QY 334 TGTGGGCGGAGCGGCTATCCATTGTCCTGCGGGAAGTGTGTTGGTGAGGCAATTGC 275  
DB |||||  
QY 2101 CCGAGCGCTTGGTGGGACTTCTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 2160  
DB |||||  
QY 274 CCGAGCGCTTGGTGGGACTTCTCCGAAACCAAGCTGCTCCCATGCTAAAGCAGG 215  
DB |||||  
QY 2161 AGAAGTGATCTCTCGATGAGATACCATTTGACACACCGATGATTTGGCAAGTGGG 2220  
DB |||||  
QY 214 AGAAGTGATCTCTCGATGAGATACCATTTGACACACCGATGATTTGGCAAGTGGG 155  
DB |||||  
QY 2221 CBTGAATCTAGATCTCTAGTAGACTCAGACGCGCTGTTGATGCAAGCAATCGAGG 2280  
DB |||||  
QY 154 CBTGAATCTAGATCTCTAGTAGACTCAGACGCGCTGTTGATGCAAGCAATCGAGG 95  
DB |||||  
QY 2281 ATTGGGCGCTTGTACTTCTGAAAGGTTTCAAGGTTTTCATCTTCCGCGGAGGAA 2340  
DB |||||  
QY 94 ATTGGGCGCTTGTACTTCTGAAAGGTTTCAAGGTTTTCATCTTCCGCGGAGGAA 35  
DB |||||  
QY 2341 TTGGGCGGAGGAGTAACACCTTTCAGCAATGG 2374  
DB |||||  
QY 34 TTGGGCGGAGGAGTAACACCTTTCAGCAATGG 1

## RESULT 2

ID AAH68528  
XX  
XX AAH68528 standard; DNA; 349980 BP.

AC AAH68528;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 7063.

KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOM ) KYOMA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

PS Disclosure; SEQ ID NO: 7063; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 349980 BP; 80727 A; 91049 C; 97618 G; 80586 T; 0 other;

Query Match 100.0%; Score 2374; DB 22; Length 349980;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGATACTCTTTGGAGAAACCATGTACGATTTGGTGACATTTGCTGCGTCTGGAAGCG 60

DB 127595 AGATACTCTTTGGAGAAACCATGTACGATTTGGTGACATTTGCTGCGTCTGGAAGCG 127654

DB 127655 TCTTTACGTGGGATTTCTTCTACGCTCCAGAGCTCACAGCGAGCGGCTGAGTTCAT 127714

QY 61 TCTTTACGTGGGATTTCTTCTACGCTCCAGAGCTCACAGCGAGCGGCTGAGTTCAT 120

QY 121 GCGGAGAGGCGCTCCCGCTTCTGATTCATCAGCAAGCTATTCATCATTAATCGTTG 180

DB 127715 GCGGAGAGGCGCTCCCGCTTCTGATTCATCAGCAAGCTATTCATCATTAATCGTTG 127774

QY 181 GGTGAGGAAACCGGCGGATGACGGTGAGAACTTGTTCAGTTCAGTCCCAAAATGGTCT 240

DB 127775 GGTGAGGAAACCGGCGGATGACGGTGAGAACTTGTTCAGTTCAGTCCCAAAATGGTCT 127834

QY 241 TGGCGTCAATGCTTTCTCACCATTGCGCAGGCGCTTGCTACGGACAATAATCTCGATGG 300

DB 127835 TGGCGTCAATGCTTTCTCACCATTGCGCAGGCGCTTGCTACGGACAATAATCTCGATGG 127894

QY 301 AATTCCAGAGGTTCCCGCGCAGCAGGTTAAGTCCCTGTCTGAGGCGATGTTGAACTG 360

DB 127895 AATTCCAGAGGTTCCCGCGCAGCAGGTTAAGTCCCTGTCTGAGGCGATGTTGAACTG 127954

QY 361 GAAACAATATGATATGGTCCCAAGCTCAATGACATCGCCAGGAAACGCGGCGAGTCACT 420

DB 127955 GAAACAATATGATATGGTCCCAAGCTCAATGACATCGCCAGGAAACGCGGCGAGTCACT 128014

QY 421 TGGCGAGATGCGCTTGCTATGGGTGCTCGCGAGCAAGAGAGTACGCGCGGATACCGT 480

DB 128015 TGGCGAGATGCGCTTGCTATGGGTGCTCGCGAGCAAGAGAGTACGCGCGGATACCGT 128074

QY 481 GACCAGTCATTTGATTTGGTCTTCTGATTTGAGCAGCTGACAAACGCTTGAATCACT 540

DB 128075 GACCAGTCATTTGATTTGGTCTTCTGATTTGAGCAGCTGACAAACGCTTGAATCACT 128134

QY 541 CAACAACCTTGGAGTTTCTTGAGCGGAGTTGGAGCGGATCGATGAGATTTCCCGAGCGG 600

|||||  
Db 128135 CAACAATGGAGTTTCTGACGCGAGTTGGAGCGATCGATGAGATTTCACACGAGC 128194  
QY 601 CGGATCAACATTTGGGCGAAGCCACCGATTCACAAACCCCGGAAACCTAACCCATCAA 660  
Db 128195 CGGCATCAACATTTGGGCGAAGCCACCGATTCACAAACCCCGGAAACCTAACCCATCAA 128254  
QY 661 CATCAGTTTGATGGCCATCGGTCATCAACATGCGACGACAGCTTGATCCAGCGCA 720  
Db 128255 CATCAGTTTGATGGCCATCGGTCATCAACATGCGACGACAGCTTGATCCAGCGCA 128314  
QY 721 CACTTTGGGCTGACACAGCGGCGTGACAAATGCTGTCGCGCAAAACCCACAGCGGAA 780  
Db 128315 CACTTTGGGCTGACACAGCGGCGTGACAAATGCTGTCGCGCAAAACCCACAGCGGAA 128374  
QY 781 CCAGATCAGGCTTTGCCGCGAAGCGCGCAGCGGCGAAATCCACCGTCGGTGTCGCGTA 840  
Db 128375 CCAGATCAGGCTTTGCCGCGAAGCGCGCAGCGGCGAAATCCACCGTCGGTGTCGCGTA 128434  
QY 841 TTGGGCGCGACGCGCGGATTAACACAAACGGGTCCAAATACGATTCGGTTCAACCA 900  
Db 128435 TTGGGCGCGACGCGCGGATTAACACAAACGGGTCCAAATACGATTCGGTTCAACCA 128494  
QY 901 GGTACAGCATTTGCCATCAACATGGGCTTTACCAAACCGGTGCTTATCGACGCTCAC 960  
Db 128495 GGTACAGCATTTGCCATCAACATGGGCTTTACCAAACCGGTGCTTATCGACGCTCAC 128554  
QY 961 CTCACCCGACCGGTTGCGCGTGTGTCAGTGGCCACCGCGAAACCGCCAAAGCGGTGTC 1020  
Db 128555 CTCACCCGACCGGTTGCGCGTGTGTCAGTGGCCACCGCGAAACCGCCAAAGCGGTGTC 128614  
QY 1021 ATCGGACCGGTTGGTCTGTTCTTCAATGATCTGGCGCTTCACCTTGTGTGTCAT 1080  
Db 128615 ATCGGACCGGTTGGTCTGTTCTTCAATGATCTGGCGCTTCACCTTGTGTGTCAT 128674  
QY 1081 GGGCTCTTTGCGTGCCATGACGGCAACCATTAACAGGTAAGCGATGCCACCCAGCGCAT 1140  
Db 128675 GGGCTCTTTGCGTGCCATGACGGCAACCATTAACAGGTAAGCGATGCCACCCAGCGCAT 128734  
QY 1141 AATATGAGACGATCGCGCGCATTTGGACAAAGATCAACGCCCAAGTGC CGCGAT 1200  
Db 128735 AATATGAGACGATCGCGCGCATTTGGACAAAGATCAACGCCCAAGTGC CGCGAT 128794  
QY 1201 GAACAAAGACGTCAGAAATTAACACACAGAGNACCGCAATGATCTCTCGGCTT 1260  
Db 128795 GAACAAAGACGTCAGAAATTAACACACAGAGNACCGCAATGATCTCTCGGCTT 128854  
QY 1261 AATTCCTTGTTTAATCACCAGTACATTTCTGCGTCCGATGACAGTAAAGACTTGGCCCC 1320  
Db 128855 AATTCCTTGTTTAATCACCAGTACATTTCTGCGTCCGATGACAGTAAAGACTTGGCCCC 128914  
QY 1321 CAAAGCAGACCTGTATAGAGATTTCCATGATCACCATCGTGCACCTATGGAAGTACTTA 1380  
Db 128915 CAAAGCAGACCTGTATAGAGATTTCCATGATCACCATCGTGCACCTATGGAAGTACTTA 128974  
QY 1381 AGTAAATGATGGTCTTTAATGATGGTTTAAATAGCTTATAGCTTATGNAACCCATCACTGGA 1440  
Db 128975 AGTAAATGATGGTCTTTAATGATGGTTTAAATAGCTTATGNAACCCATCACTGGA 129034  
QY 1441 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGGCTCTCTAGCCCTTTCCAT 1500  
Db 129035 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGGCTCTCTAGCCCTTTCCAT 129094  
QY 1501 TTCCCTTCGGCGGTGATGACGCGGTAAAGCTCTCGAGATCACTGCGGTGAGTGT 1560  
Db 129095 TTCCCTTCGGCGGTGATGACGCGGTAAAGCTCTCGAGATCACTGCGGTGAGTGT 129154  
QY 1561 GGTATCGGACCAACCGGCCAAGCAACGAGCGGCTGAGTCTCTGCAAGCAGC 1620  
Db 129155 GGTATCGGACCAACCGGCCAAGCAACGAGCGGCTGAGTCTCTGCAAGCAGC 129214  
QY 1621 GCGGAAATGGTGTGTCGCAAGCAGAACTTAAAGCGCAACTATCTGAGCGCTTGTCTGA 1680  
|||||

Db 129215 GCGGAAATGGTGTGTCGCAAGCAGAAACTAAAGCGCAACTATCTGAGCGCTTGTCTGA 129274  
QY 1681 AATCCGTTAAACATCGCCATCAACGAGATTCGTATATCCATGTTTCTCCCGTGT 1740  
Db 129275 AATCCGTTAAACATCGCCATCAACGAGATTCGTATATCCATGTTTCTCCCGTGT 129334  
QY 1741 CAAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTGCGCTTTGGAAGATGAAGCGCA 1800  
Db 129335 CAAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCTGCGCTTTGGAAGATGAAGCGCA 129394  
QY 1801 CACATTTATCTTCTGCTGCGCGCTGGAGATGTTTATGAGAGCGGTAAACCGTGAAGCTAATCC 1860  
Db 129395 CACATTTATCTTCTGCTGCGCGCTGGAGATGTTTATGAGAGCGGTAAACCGTGAAGCTAATCC 129454  
QY 1861 CTTGCGGAGATGTAAGTAGTAGTAAGCTTGGAACCATGCGCCACTTGGCCATTTGCAACCCC 1920  
Db 129455 CTTGCGGAGATGTAAGTAGTAGTAAGCTTGGAACCATGCGCCACTTGGCCATTTGCAACCCC 129514  
QY 1921 CTCATTGGGGATGCTTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGTCTT 1980  
Db 129515 CTCATTGGGGATGCTTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGTCTT 129574  
QY 1981 ACGCTTCGGTCCCAAGATGCTCTTCAAGACCGTGCACCTGGACGGCGCTCGATGGTCC 2040  
Db 129575 ACGCTTCGGTCCCAAGATGCTCTTCAAGACCGTGCACCTGGACGGCGCTCGATGGTCC 129634  
QY 2041 TGTGGGCGCAGCGCGGTATTCATTGTCCTGCGCGGAGAGTTTGTGTGAGGCAATTCG 2100  
Db 129635 TGTGGGCGCAGCGCGGTATTCATTGTCCTGCGCGGAGAGTTTGTGTGAGGCAATTCG 129694  
QY 2101 CCGAGGCTTGGTTCGGGACTTCTTCCGGAACCAAGCTGCTCCCATGCTAAAGCAGG 2160  
Db 129695 CCGAGGCTTGGTTCGGGACTTCTTCCGGAACCAAGCTGCTCCCATGCTAAAGCAGG 129754  
QY 2161 AGAAGTGATCTCTCGATGAGATACCCATTGACACACCGATGTATGGAACGATGGCG 2220  
Db 129755 AGAAGTGATCTCTCGATGAGATACCCATTGACACACCGATGTATGGAACGATGGCG 129814  
QY 2221 CTTGAGATCTAGATCTCTAGCTAGACTACACACCGCTGCTGTGATGACGAAATCGAGG 2280  
Db 129815 CTTGAGATCTAGATCTCTAGCTAGACTACACACCGCTGCTGTGATGACGAAATCGAGG 129874  
QY 2281 ATTGGCGCTTACTTCTGAAAAGTTTCAAGGTTTTTCACTTCTTCGCGCGCAGGAA 2340  
Db 129875 ATTGGCGCTTACTTCTGAAAAGTTTCAAGGTTTTTCACTTCTTCGCGCGCAGGAA 129934  
QY 2341 TTGGGCGCAGCAGAGTAACACCTTCAGCAATGG 2374  
Db 129935 TTGGGCGCAGCAGAGTAACACCTTCAGCAATGG 129968  
|||||

## RESULT 3

AAF71777

ID AAF71777 standard, DNA; 993 BP.

XX.

AC AAF71777;

XX

DT 30-APR-2001 (first entry)

XX

Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:49.

XX

Corynebacterium glutamicum; metabolic pathway protein; MP protein;

fine chemical production; microorganism; organic acid; nucleoside;

nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;

lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;

carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.

XX

Corynebacterium glutamicum.

OS

WO200100843-A2.

XX

PD 04-JAN-2001.

XX



```

Db 841 AGAAGTGATCCTCTCGATGAGATACCAATGACACACCAGATGATGGCAAGTGCGG 900
QY 2221 CCTGGAACTAGATCTCTAGCTAGACTCACAGACGCCGCTGTTGATGACGAATCGAGGG 2280
Db 901 CCTGGAACTAGATCTCTAGCTAGACTCACAGACGCCGCTGTTGATGACGAATCGAGGG 960
QY 2281 ATTGGCGCCTTAGTTACTTCTGAAAAGGTTTCAG 2313
Db 961 ATTGGCGCCTTAGTTACTTCTGAAAAGGTTTCAG 993

RESULT 4
AAS96096
ID AAS96096 standard; DNA; 993 BP.
XX
AC AAS96096;
XX
DT 26-FEB-2002 (first entry)
XX
DE C. glutamicum gene #21 encoding metabolic pathway protein.
XX
KW Metabolic pathway protein; MP; lysine biosynthesis pathway;
KW methionine biosynthesis pathway; large-scale production of fine chemical;
KW Corynebacterium diphtheriae; diphtheria; ds.
XX
OS Corynebacterium glutamicum.
XX
PN W0200166573-A2.
XX
PD 13-SEP-2001.
XX
PF 22-DEC-2000; 2000WO-1B02035.
XX
PR 09-MAR-2000; 2000US-187970P.
XX
PR 23-JUN-2000; 2000US-0606740.
XX
PA (BADI ) BASF AG.
XX
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;
PI Lee H, Hwang B;
XX
DR WPI; 2001-582269/65.
DR P-PSDB; AAU71886.
XX
XX Nucleic acids encoding metabolic pathway proteins from Corynebacterium
PT glutamicum, useful for producing methionine and lysine in
PT Corynebacterium and Brevibacterium.
XX
PS Disclosure; Page 207-208; 316pp; English.
XX
CC The present invention relates to the isolation of novel Corynebacterium
CC glutamicum genes encoding metabolic pathway (MP) proteins
CC (AAU71863-AAU71922). The metabolic pathway proteins of the invention
CC include enzymes involved in the lysine and methionine biosynthetic
CC pathways. The polynucleotide sequences of the invention can be used
CC for the large-scale production and/or modulation of expression of the
CC fine chemicals such as lysine and methionine. The sequences of the
CC invention may be used to identify C. glutamicum and related organisms
CC e.g. C. diphtheriae in a subject to detect diphtheria.
CC AAS96073-AAS96132 represent C. glutamicum genes encoding the novel
CC metabolic pathway proteins of the invention.
XX
SQ Sequence 993 BP; 233 A; 256 C; 268 G; 236 T; 0 other;

Query Match 41.8%; Score 993; DB 23; Length 993;
Best Local Similarity 100.0%; Pred. No. 2.6e-299;
Matches 993; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1321 CAAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 1380
Db 1 CAAAAGCAGACCTGTAATGAAGATTTCCATGATCACCATCGTGACCTATGGAAGTACTTA 60

```

```

QY 1381 AGTAAATGATTTGGTCTTAAACATGGTTTAAATATAGCTTCATGAACCCCAATTCACATCGA 1440
Db 61 AGTAAATGATTTGGTCTTAAACATGGTTTAAATATAGCTTCATGAACCCCAATTCACATCGA 120
QY 1441 CACTTGTCTCTCAATCAATGATGAAGGAGCTTCGAAGGCGCTCTTGAAGCCCTTTCCAT 1500
Db 121 CACTTGTCTCTCAATCAATGATGAAGGAGCTTCGAAGGCGCTCTTGAAGCCCTTTCCAT 180
QY 1501 TTCCCCCTCGGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACTGCGGTGAGTGT 1560
Db 181 TTCCCCCTCGGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCACTGCGGTGAGTGT 240
QY 1561 GGTATCGGCACCCCAACCGGCCAAAGCAACCGAGGGGTGAAGTCTTGTGCAAGCAGC 1620
Db 241 GGTATCGGCACCCCAACCGGCCAAAGCAACCGAGGGGTGAAGTCTTGTGCAAGCAGC 300
QY 1621 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGGACGCTTGTCTGA 1680
Db 301 GCGGAAATGGTGTGCTGCAAGCAGAACTAAAGCGCAACTATCTGGACGCTTGTCTGA 360
QY 1681 AATCCGTTAAACATCGCCATCAACGAGATTCGCTATCCATGATGTTTCTCCCGTGT 1740
Db 361 AATCCGTTAAACATCGCCATCAACGAGATTCGCTATCCATGATGTTTCTCCCGTGT 420
QY 1741 CAACGAGGTAGCTTCTTGGGTTGGAGCAACGCTCAGCTCGCTTGGAGATGAAGCGCA 1800
Db 421 CAACGAGGTAGCTTCTTGGGTTGGAGCAACGCTCAGCTCGCTTGGAGATGAAGCGCA 480
QY 1801 CACATTATCTTGTGCGCGGTGGAGATGTTTGTAGAGCGGTAAACCGGTGAAGCTAATCC 1860
Db 481 CACATTATCTTGTGCGCGGTGGAGATGTTTGTAGAGCGGTAAACCGGTGAAGCTAATCC 540
QY 1861 CGTGGCGGATGTAAGTAGTAGAACTTGGAAACCATGCGCCACTTGGCCATTTGCAACCCC 1920
Db 541 CGTGGCGGATGTAAGTAGTAGAACTTGGAAACCATGCGCCACTTGGCCATTTGCAACCCC 600
QY 1921 CTCATTGCGGGATGCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTTT 1980
Db 601 CTCATTGCGGGATGCTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTTT 660
QY 1981 AGCTTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGACCTGGACGGGCGGTGATGGTCC 2040
Db 661 AGCTTTCGGTCCCAAGATGTCTTCAAGACCGTGAACCTGACCTGGACGGGCGGTGATGGTCC 720
QY 2041 TGTGGGCGCAGCGCGTATCCATGTCCTGCGCGGAAGGTTTGTGAGGCAATTCG 2100
Db 721 TGTGGGCGCAGCGCGTATCCATGTCCTGCGCGGAAGGTTTGTGAGGCAATTCG 780
QY 2101 CCGAGGCTTGGTGGGGACTTCTTCCGAAACCCCAAGCTGCTCCCATGCTAAAGCAGG 2160
Db 781 CCGAGGCTTGGTGGGGACTTCTTCCGAAACCCCAAGCTGCTCCCATGCTAAAGCAGG 840
QY 2161 AGAAGTGATCTCTCGATGAGATACCCATTGACACACCGATGATTTGGCAACGATGGCG 2220
Db 841 AGAAGTGATCTCTCGATGAGATACCCATTGACACACCGATGATTTGGCAACGATGGCG 900
QY 2221 CTGGAATCTAGATCTCTAGCTAGACTCACAGACGCCGCTGTTGATGACGAATTCAGGG 2280
Db 901 CTGGAATCTAGATCTCTAGCTAGACTCACAGACGCCGCTGTTGATGACGAATTCAGGG 960
QY 2281 ATTGGCGCCTTAGTTACTTCTGAAAAGGTTTCAG 2313
Db 961 ATTGGCGCCTTAGTTACTTCTGAAAAGGTTTCAG 993

RESULT 5
AAH68421
ID AAH68421 standard; DNA; 870 BP.
XX
AC AAH68421;
XX
DT 26-SEP-2001 (first entry)
XX

```





PR 08-JUL-1999; 99DE-1031434.  
 PR 08-JUL-1999; 99DE-1031435.  
 PR 08-JUL-1999; 99DE-1031443.  
 PR 08-JUL-1999; 99DE-1031453.  
 PR 08-JUL-1999; 99DE-1031457.  
 PR 08-JUL-1999; 99DE-1031465.  
 PR 08-JUL-1999; 99DE-1031478.  
 PR 08-JUL-1999; 99DE-1031510.  
 PR 08-JUL-1999; 99DE-1031541.  
 PR 08-JUL-1999; 99DE-1031573.  
 PR 08-JUL-1999; 99DE-1031592.  
 PR 08-JUL-1999; 99DE-1031632.  
 PR 08-JUL-1999; 99DE-1031634.  
 PR 08-JUL-1999; 99DE-1031636.  
 PR 08-JUL-1999; 99DE-1032125.  
 PR 08-JUL-1999; 99DE-1032126.  
 PR 08-JUL-1999; 99DE-1032130.  
 PR 08-JUL-1999; 99DE-1032186.  
 PR 08-JUL-1999; 99DE-1032206.  
 PR 08-JUL-1999; 99DE-1032227.  
 PR 08-JUL-1999; 99DE-1032228.  
 PR 08-JUL-1999; 99DE-1032229.  
 PR 08-JUL-1999; 99DE-1032230.  
 PR 14-JUL-1999; 99DE-1032922.  
 PR 14-JUL-1999; 99DE-1032926.  
 PR 14-JUL-1999; 99DE-1032928.  
 PR 14-JUL-1999; 99DE-1033004.  
 PR 14-JUL-1999; 99DE-1033005.  
 PR 14-JUL-1999; 99DE-1033006.  
 PR 12-AUG-1999; 99US-0148613.  
 PR 27-AUG-1999; 99DE-1040764.  
 PR 27-AUG-1999; 99DE-1040765.  
 PR 27-AUG-1999; 99DE-1040766.  
 PR 27-AUG-1999; 99DE-1040832.  
 PR 31-AUG-1999; 99DE-1041378.  
 PR 31-AUG-1999; 99DE-1041379.  
 PR 31-AUG-1999; 99DE-1041380.  
 PR 31-AUG-1999; 99DE-1041394.  
 PR 31-AUG-1999; 99DE-1041396.  
 PR 03-SEP-1999; 99DE-1042076.  
 PR 03-SEP-1999; 99DE-1042077.  
 PR 03-SEP-1999; 99DE-1042079.  
 PR 03-SEP-1999; 99DE-1042086.  
 PR 03-SEP-1999; 99DE-1042087.  
 PR 03-SEP-1999; 99DE-1042088.  
 PR 03-SEP-1999; 99DE-1042095.  
 PR 03-SEP-1999; 99DE-1042124.  
 PR 03-SEP-1999; 99DE-1042129.  
 PR 09-MAR-2000; 2000US-0187970.  
 XX (BADI ) BASF AG.  
 PA Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauser G;  
 PI WPI; 2001-137957/14.  
 XX P-PSDB; AAB79660.  
 DR Nucleic acids from *Corynebacterium glutamicum* encoding metabolic  
 PT pathway proteins, useful for producing fine chemicals in  
 PT microorganisms, including organic acids, nonproteinogenic amino acids,  
 PT and purine and pyrimidine bases -  
 XX Claim 3; Page 233-234; 1737pp; English.  
 PS CC AAF71753 to AAF72330 encode the *Corynebacterium glutamicum* metabolic  
 CC pathway (MP) proteins given in AAB79634 to AAB80211. The *C. glutamicum*  
 CC MP nucleic acids are useful for the production of fine chemicals  
 CC in microorganisms, including organic acids, nonproteinogenic amino  
 CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,  
 CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic  
 CC compounds, vitamins, cofactors, polyketides and enzymes.  
 XX Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;  
 SQ

Query Match 34.6%; Score 822; DB 22; Length 822;  
 Best Local Similarity 100.0%; Pred. No. 6.5e-246;  
 Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 629 GATTCACAAACCCGGGAAACTAACCCATCAACATCAGTTTTCATGGCCCAATCGGTCATC 688  
 DB 822 GATTCACAAACCCGGGAAACTAACCCATCAACATCAGTTTTCATGGCCCAATCGGTCATC 763  
 QY 689 ACAACTGCCACGACGAGTTCATCCAGCCCAACACTTGGGGCTGGACAGCGGGCGTGCAC 748  
 DB 762 ACAACTGCCACGACGAGTTCATCCAGCCCAACACTTGGGGCTGGACAGCGGGCGTGCAC 703  
 QY 749 AATGCTGCTGCCGCGGAAACCCGACGAGCGGGAACAGATCAGGCTTGCAGCAACCGGCA 808  
 DB 702 AATGCTGCTGCCGCGGAAACCCGACGAGCGGGAACAGATCAGGCTTGCAGCAACCGGCA 643  
 QY 809 GCGGCGAAATCCACCGTCCGGTGTTCGGCGTATTCGGCCGACGCGCGCGGATTAACACA 868  
 DB 642 GCGGCGAAATCCACCGTCCGGTGTTCGGCGTATTCGGCCGACGCGCGCGGATTAACACA 583  
 QY 869 AACGGCTCCAAATACGCAATTCGGGTTCAACAGGTTCAGCAGCATTCGCCATCAACATGGGC 928  
 DB 582 AACGGCTCCAAATACGCAATTCGGGTTCAACAGGTTCAGCAGCATTCGCCATCAACATGGGC 523  
 QY 929 TTATCCCAAAACCGGCTGCTTATCGAGCTCAGCTCCACCGCACCGGTTGCGCGTGTCA 988  
 DB 522 TTATCCCAAAACCGGCTGCTTATCGAGCTCAGCTCCACCGCACCGGTTGCGCGTGTCA 463  
 QY 989 GTGGCCACCGCGCAACCGCCCAAGGCGTGTATCGGGCAGCGGTTGTTCTGTTCTTCA 1048  
 DB 462 GTGGCCACCGCGCAACCGCCCAAGGCGTGTATCGGGCAGCGGTTGTTCTGTTCTTCA 403  
 QY 1049 ATGATCTGTGGCGCTTCCACCTTGTGTGTCATGGGCTCTTCGCTGCCATGACGCAAC 1108  
 DB 402 ATGATCTGTGGCGCTTCCACCTTGTGTGTCATGGGCTCTTCGCTGCCATGACGCAAC 343  
 QY 1109 CATAACAGGTAAGCGATGCGCCAGCGCATTAATTCAGCAGCATTCGCGCGGCGGATTCG 1168  
 DB 342 CATAACAGGTAAGCGATGCGCCAGCGCATTAATTCAGCAGCATTCGCGCGGCGGATTCG 283  
 QY 1169 GACAAAAGATCAACGCCCAAGGTGCGCGGATGAACAAAGACGTCAGAAATTAACAC 1228  
 DB 282 GACAAAAGATCAACGCCCAAGGTGCGCGGATGAACAAAGACGTCAGAAATTAACAC 223  
 QY 1229 ACAGAAAGAACCGCAATGAGTCTTCGCGCTTAATTCCTTGTGTTAATCACCAGTACATTC 1288  
 DB 222 ACAGAAAGAACCGCAATGAGTCTTCGCGCTTAATTCCTTGTGTTAATCACCAGTACATTC 163  
 QY 1289 TCGGCTCCGATGGACAGTAAGAGATGCGCCCGCAAGCAGACCTGTAATGAGATTTC 1348  
 DB 162 TCGGCTCCGATGGACAGTAAGAGATGCGCCCGCAAGCAGACCTGTAATGAGATTTC 103  
 QY 1349 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATTGTTCTTAACATGGTT 1408  
 DB 102 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATTGTTCTTAACATGGTT 43  
 QY 1409 TAATATAGCTTCATGAACCCCATTCACCTGGACACTTTGCTC 1450  
 DB 42 TAATATAGCTTCATGAACCCCATTCACCTGGACACTTTGCTC 1

## RESULT 7

AAS96098/c

ID AAS96098 standard; DNA; 822 BP.

XX AC AAS96098;

XX DT 26-FEB-2002 (first entry)

XX DE C. glutamicum gene #23 encoding metabolic pathway protein.

XX KW Metabolic pathway protein; MP; lysine biosynthesis pathway;

methionine biosynthesis pathway; large-scale production of fine chemical;  
Corynebacterium diphtheriae; diphtheria; ds.

Corynebacterium glutamicum.

W0200166573-A2.

13-SEP-2001.

22-DEC-2000; 2000WO-IB02035.

09-MAR-2000; 2000US-187970P.

23-JUN-2000; 2000US-0606740.

(BADI ) BASF AG.

Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G, Kim J;  
Lee H, Hwang B;

WPI; 2001-582269/65.

P-PSDB; AAU71888.

Nucleic acids encoding metabolic pathway proteins from Corynebacterium  
glutamicum, useful for producing methionine and lysine in  
Corynebacterium and Brevibacterium -

Disclosure; Page 214-215; 316pp; English.

The present invention relates to the isolation of novel Corynebacterium  
glutamicum genes encoding metabolic pathway (MP) proteins  
(AAU71863-AAU71922). The metabolic pathway proteins of the invention  
include enzymes involved in the lysine and methionine biosynthetic  
pathways. The polynucleotide sequences of the invention can be used  
for the large-scale production and/or modulation of expression of  
fine chemicals such as lysine and methionine. The sequences of the  
invention may be used to identify C. glutamicum and related organisms  
e.g. C. diphtheriae in a subject to detect diphtheria.  
AAU96073-AAU96132 represent C. glutamicum genes encoding the novel  
metabolic pathway proteins of the invention.

Sequence 822 BP; 167 A; 192 C; 246 G; 217 T; 0 other;

Query Match 34.6%; Score 822; DB 23; Length 822;

Best Local Similarity 100.0%; Pred. No. 6.5e-246;

Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCCAAACCCGCGAAACTAACCCATCAACATCAGTTGATGGCCATGCGGTATC 688

DB 822 GATTCCAAACCCGCGAAACTAACCCATCAACATCAGTTGATGGCCATGCGGTATC 763

QY 689 ACAACTGCCAGCAGCGAGTTGATCCAGCGCCACACCTTGGGGCTGGACAGCGCGGTGAC 748

DB 762 ACAACTGCCAGCAGCGAGTTGATCCAGCGCCACACCTTGGGGCTGGACAGCGCGGTGAC 703

QY 749 AATGCTGCTGCGCGAAACCCACAGCGGAAACAGATCAGGCTTGCAGCAACGCGCCA 808

DB 702 AATGCTGCTGCGCGAAACCCACAGCGGAAACAGATCAGGCTTGCAGCAACGCGCCA 643

QY 809 GCGCGGAAATCCACCGTCCGGTGTCCCGCTATTCGCGCGAGCGCGCGATTAACACA 868

DB 642 GCGCGGAAATCCACCGTCCGGTGTCCCGCTATTCGCGCGAGCGCGCGATTAACACA 583

QY 869 AAGCGCTCCAAATACGATTCGGGTTCAACAGGTGACAGCAGTTCATCAACATGGGC 928

DB 582 AAGCGTCCAAATACGATTCGGGTTCAACAGGTGACAGCAGTTCATCAACATGGGC 523

QY 929 TTTACCAAAACCCGCTGTTATCGAGCTCACCTCCACCGCGGTGCGCGGTGCA 988

DB 522 TTTACCAAAACCCGCTGTTATCGAGCTCACCTCCACCGCGGTGCGCGGTGCA 463

QY 989 GTGGCCACCGCGCAACCGCCAAAGGGGTGTCATCGGGCAAGTGTGTTCTTCA 1048

DB 462 GTGGCCACCGCGCAACCGCCAAAGGGGTGTCATCGGGCAAGTGTGTTCTTCA 403

QY 1049 ATGATCTGTGGCGCTTCCACCTTGTTGTTCATGGCGTCTTTCCTGCCATGACGCGCAAC 1108

DB 402 ATGATCTGTGGCGCTTCCACCTTGTTGTTCATGGCGTCTTTCCTGCCATGACGCGCAAC 343

QY 1109 CATAACAGGTAAAGCGATGCCACCCAGCGCATATAATCGAGCACGATCGCGCGCATTTG 1168

DB 342 CATAACAGGTAAAGCGATGCCACCCAGCGCATATAATCGAGCACGATCGCGCGCATTTG 283

QY 1169 GACAAAAGATCAACGCCCAAGGTGCGCGCGATGAACAAAAGACGTCAGAAAATTAACAC 1228

DB 282 GACAAAAGATCAACGCCCAAGGTGCGCGCGATGAACAAAAGACGTCAGAAAATTAACAC 223

QY 1229 ACGAGAAGAACCGCAATGAGTCTTCGCGCTTAATTCCTGTTTATCACCAGTACATTC 1288

DB 222 ACGAGAAGAACCGCAATGAGTCTTCGCGCTTAATTCCTGTTTATCACCAGTACATTC 163

QY 1289 TGGCGTCCGATGGACAGTAAAGACTGGCCCCCAAAAGCAGACCTGTAATGAAGATTTC 1348

DB 162 TGGCGTCCGATGGACAGTAAAGACTGGCCCCCAAAAGCAGACCTGTAATGAAGATTTC 103

QY 1349 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATTTGTTCTTAAACATGGTT 1408

DB 102 ATGATCACCATCGTGACCTATGGAAGTACTTAAGTAAATGATTTGTTCTTAAACATGGTT 43

QY 1409 TAATATAGCTTCATGAACCCCATTCACCTGGACACTTTTGCTC 1450

DB 42 TAATATAGCTTCATGAACCCCATTCACCTGGACACTTTTGCTC 1

# RESULT 8

AAH68420/c

ID AAH68420 standard; DNA; 708 BP.

AC AAH68420;

DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 3455.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX organic acid synthesis; ds.

OS Corynebacterium glutamicum.

PN EP1108790-A2.

PD 20-JUN-2001.

PF 18-DEC-2000; 2000EP-0127688.

PR 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

PA (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

DR P-PSDB; AAG93201.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

PS Claim 1; SEQ ID NO: 3455; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a





QY 1190 GTGCGCGATGAACAAAGACGCTCAGAAATTAACACACAGAGAAGAACCGCAATGAGT 1249  
| | | | |  
Db 947 GTGCGGAGGTGAACAGGACACGCTGCGACAGACAGATGATGACGCGCGTGTATG 888  
| | | | |  
QY 1250 CTTTCGGCTTAATTCCTTGTGTTAATCACCAGTACATTCGCGTCCGATGAGCAGTAAA 1309  
| | | | |  
Db 887 CCCTCGGCTTGTATGCGCTGTGTTGATCACCAGGACATTCGTGCGCGCGATGCGCAACAGC 828  
| | | | |  
QY 1310 AGACTGCGCCCAAAAGCAGACCTGTATGAAGATTTCATGATCACCATCGTGACCTAT 1369  
| | | | |  
Db 827 AGACTGCTCCCAACACAAACCGGTGCAAGATTTCATGTCGCGATGTTCCGCGAC 768  
| | | | |  
QY 1370 GGAAGTACTTAAGTAAATGATGGTCTTAATGATGTTTAAATATAGCTTCATGAACCCC 1429  
| | | | |  
Db 767 AAACACCATTAAGTAAATGATAGAGACTGAACATGTTTGAAGATTGCTTCATGAACCCG 708  
| | | | |  
QY 1430 ATTCAATGAGACCTTGTCTCTCAATCATGATGAAGCGAGCTTCGAAGGCGCTCTTA 1489  
| | | | |  
Db 707 ATCCACTGAGACCCCTTCTGACCATCATCGACGAGGCGAGCTTCGAGAACGCCCTCCCTC 648  
| | | | |  
QY 1490 GCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGTG 1549  
| | | | |  
Db 647 GCATGTGATCTCCCTCGCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGTG 588  
| | | | |  
QY 1550 GGTGAGTGTGATCGCGCACCAACCGGCCAAAGCGGCTGAAGTCTCTT 1609  
| | | | |  
Db 587 GGTGCGGTACTGTGTGCGCACCGCGGTGCGCGGTTGCCACGAAAGCGGGAGTGTG 528  
| | | | |  
QY 1610 GTGCAAGCAGCGGAAATGTTGCTGTCAGCAGCAAACTAAAGCGCAACTATCTGGA 1669  
| | | | |  
Db 527 GTCCAGCGCGCGCAAGATGGCGCTCTGCGAGCGGAGACCCGTGAGCAACTGGCGGAA 468  
| | | | |  
QY 1670 CGCTTGTGTAATCCCGTTAAACCATCCATCAACGAGATCGCTATCCACATGTTT 1729  
| | | | |  
Db 467 CGGTGACGAGATCCCTGACCGTGTGCTATCAACGCGATTCCTGTCCACCTGTTT 408  
| | | | |  
QY 1730 CTTCCGCTGTTCAAGAGGTAGTCTTGTGGGTGAGCAACGCTCAGCTGCGCTTGGA 1789  
| | | | |  
Db 407 CGCGCGTGTTCGCGAGGTGACATTTGGGTGCGGTGCGCTGACCTGACGCTGGTGGAG 348  
| | | | |  
QY 1790 GATCAAGCGCACATATCTCTGCTGCGGTGAGATGTTTGAAGCGGTAAACCGT 1849  
| | | | |  
Db 347 GACGAGCCCAACACCTGTCTCTGCTGCGCGGTTCGCTGCGGCGCGTGAACCGC 288  
| | | | |  
QY 1850 GAAGCTAATCCGCGGCGATGTAAGTAGTAACTTGAACCATGCGCACTTGCGC 1909  
| | | | |  
Db 287 GAGCGGACCCGTTGCGGGTGTAGGTCTACGCTCGGGGTATGAGACACCTTGCG 228  
| | | | |  
QY 1910 ATTGCAACCCCTCATTTGCGGGATGCTACATGTTGATGGAAACTAGATTGGGCTGCG 1969  
| | | | |  
Db 227 GTGCGCAACCCCGAATCTGAGGCAAGGTACACGTTGATGGCAACCCGACTTGGTGG 168  
| | | | |  
QY 1970 -ATGCGCGTCTTACGCTTCCGTTCCAAAGATGTCTTCAAGCGGTGACCTGAAGCGGCG 2028  
| | | | |  
Db 167 CATGCGGCTACTGCGTTTCGCGCGCAATGATGTGCTGCAGACCGCGATCTCGAGGCGAG 108  
| | | | |  
QY 2029 CGTGCATGTCTGTTGGGCGCGAGCGCTATCCATTTGCTCGCGCGCAAGGTTTGG 2088  
| | | | |  
Db 107 GGTGACGAGCGGTGCGCGCGCGCGCGCTGTCTCGGTGTCGCGGAGGTTTGG 48  
| | | | |  
QY 2089 TGAGGCAATTTGCGCGAGCGCTTGTGTTGGGCACTTCTTCCCGAAACCC 2135  
| | | | |  
Db 47 GGAGCGGTCGCACTCGCGCTGGGCTGGGCTGCTCTCCCGAGGCC 1

RESULT 10  
AAH68419  
ID AAH68419 standard; DNA; 627 BP.  
XX  
AC AAH68419;  
XX  
DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 3454.  
DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
XX organic acid synthesis; ds.  
KW Corynebacterium glutamicum.  
XX EP1108790-A2.  
PN 20-JUN-2001.  
PD 18-DEC-2000; 2000EP-0127688.  
XX 16-DEC-1999; 99JP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
PR 03-AUG-2000; 2000JP-0280988.  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
PI WPI; 2001-376931/40.  
DR P-PSDB; AAG93200.  
XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
PS Claim 1; SEQ ID NO: 3454; 246bp + Sequence Listing; English.  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
SQ Sequence 627 BP; 139 A; 159 C; 185 G; 144 T; 0 other;  
Query Match 26.4%; Score 627; DB 22; Length 627;  
Best Local Similarity 100.0%; Pred. No. 5e-185;  
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 ATGTACGCAATTCGTCGACATTTGTCGCTCGGAAAGGCTCTTTAGTGGTATTTCTTCC 82  
| | | | |  
Db 1 ATGTACGCAATTCGTCGACATTTGTCGCTCGGAAAGGCTCTTTAGTGGTATTTCTTCC 60  
| | | | |  
QY 83 TAGGTTCAGAGCTCAGCGGAGCGGCTGAGTTCATGCGGAGAGAGGCTGCGCGCTT 142  
| | | | |  
Db 61 TAGGTTCAGAGCTCAGCGGAGCGGCTGAGTTCATGCGGAGAGAGGCTGCGCGCTT 120  
| | | | |  
QY 143 CTGATTTCATCAGCAAGCTATTCATTAATCGTTGGTGGAGGACCGGCGCATGAC 202  
| | | | |  
Db 121 CTGATTTCATCAGCAAGCTATTCATTAATCGTTGGTGGAGGACCGGCGCATGAC 180  
| | | | |  
QY 203 GGTGAGAACTTGTTCAGTCAGTCGCAACATGGTCTTGGCGTCATTGCTTTCTCACCA 262  
| | | | |  
Db 181 GGTGAGAACTTGTTCAGTCAGTCGCAACATGGTCTTGGCGTCATTGCTTTCTCACCA 240  
| | | | |  
QY 263 CTTGCGAGGCGCTGCTCAGCGACAAATATCTCGATGGAATTCAGAGGTTTCCCGCGCC 322  
| | | | |  
Db 241 CTTGCGAGGCGCTGCTCAGCGACAAATATCTCGATGGAATTCAGAGGTTTCCCGCGCC 300  
| | | | |

QY 323 AGCCAGGGTAAGTCCCTGCTCTGAGGGGATGTTGAACGTGAACAATATTGATGTCCTCCG 382  
DB 301 AGCCAGGGTAAGTCCCTGCTCTGAGGGGATGTTGAACGTGAACAATATTGATGTCCTCCG 360  
QY 383 AAGCTCAATGACATCGCCAGGAAACGGGGCAGTCACTTCGCGAGATGGCGCTTGCATGG 442  
DB 361 AAGCTCAATGACATCGCCAGGAAACGGGGCAGTCACTTCGCGAGATGGCGCTTGCATGG 420  
QY 443 GTGCTGCGGAGCAAGAGAGATGACGGCGCGATACCGTGACCAAGTGCATGATGTTGCT 502  
DB 421 GTGCTGCGGAGCAAGAGAGATGACGGCGCGATACCGTGACCAAGTGCATGATGTTGCT 480  
QY 503 TCGTCAGTTGAGCAGCTGGGCAACAGCGCTTGATTTCACTCAACACTTGGATTTCTGAC 562  
DB 481 TCGTCAGTTGAGCAGCTGGGCAACAGCGCTTGATTTCACTCAACACTTGGATTTCTGAC 540  
QY 563 GCCGAGTTGAGGCGATCGATGATTTCCACGACCGCGCATCAACATTTGGCGGAAG 622  
DB 541 GCCGAGTTGAGGCGATCGATGATTTCCACGACCGCGCATCAACATTTGGCGGAAG 600  
QY 623 GCCACGATTCGAAACCCGCGAAAC 649  
DB 601 GCCACGATTCGAAACCCGCGAAAC 627

RESULT 11

AAF68077  
ID AAF68077 standard; DNA; 993 BP.  
XX AAF68077;  
XX  
DT 11-APR-2001 (first entry)  
XX  
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:669.  
XX  
KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;  
KW membrane construction and membrane transport protein; petroleum spill;  
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;  
KW identification; microorganism; fine chemical production; transformation;  
KW genome mapping; genetic engineering; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
XX WC200100805-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-IB00926.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031454.  
PR 08-JUL-1999; 99DE-1031478.  
PR 08-JUL-1999; 99DE-1031563.  
PR 08-JUL-1999; 99DE-1032122.  
PR 08-JUL-1999; 99DE-1032124.  
PR 08-JUL-1999; 99DE-1032125.  
PR 08-JUL-1999; 99DE-1032128.  
PR 08-JUL-1999; 99DE-1032180.  
PR 08-JUL-1999; 99DE-1032182.  
PR 08-JUL-1999; 99DE-1032190.  
PR 08-JUL-1999; 99DE-1032191.  
PR 08-JUL-1999; 99DE-1032209.  
PR 08-JUL-1999; 99DE-1032212.  
PR 08-JUL-1999; 99DE-1032227.  
PR 08-JUL-1999; 99DE-1032228.  
PR 08-JUL-1999; 99DE-1032229.  
PR 08-JUL-1999; 99DE-1032230.  
PR 14-JUL-1999; 99DE-1032927.  
PR 14-JUL-1999; 99DE-1033005.  
PR 14-JUL-1999; 99DE-1033006.  
PR 27-AUG-1999; 99DE-1040764.  
PR 27-AUG-1999; 99DE-1040765.  
PR 27-AUG-1999; 99DE-1040766.

PR 27-AUG-1999; 99DE-1040830.  
PR 27-AUG-1999; 99DE-1040831.  
PR 27-AUG-1999; 99DE-1040832.  
PR 27-AUG-1999; 99DE-1040833.  
PR 31-AUG-1999; 99DE-1041378.  
PR 31-AUG-1999; 99DE-1041379.  
PR 31-AUG-1999; 99DE-1041395.  
PR 03-SEP-1999; 99DE-1042077.  
PR 03-SEP-1999; 99DE-1042078.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042088.  
XX (BADI ) BASF AG.  
PA  
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Habershauer G;  
PI  
XX WPI; 2001-071486/08.  
DR P-PSDB; AAB76844.  
XX  
PT Corynebacterium glutamicum nucleic acids encoding membrane construction  
PT and membrane transport proteins or their portions, useful for typing or  
PT identifying C. glutamicum or related bacteria, and as markers for  
PT transformation  
XX  
PS Claim 3; Page 1107-1108; 1119pp; English.  
XX  
CC AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane  
CC construction and membrane transport (MCT) proteins given in AAB76510 to  
CC AAB76847. The MCT nucleic acids and proteins are useful in the  
CC identification of microorganisms which can be used to produce fine  
CC chemicals, for modulating fine chemical production in C. glutamicum or  
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or  
CC identification of C. glutamicum or related bacteria, as reference points  
CC for mapping C. glutamicum genome, and as markers for transformation.  
CC AAF68082 and AAF68082 represent sequencing primers which are used in an  
CC example from the present invention.  
XX  
SQ Sequence 993 BP; 222 A; 247 C; 277 G; 247 T; 0 other;  
XX  
Query Match 20.4%; Score 485.2; DB 22; Length 993;  
Best Local Similarity 99.0%; Pred. No. 1.4e-140;  
Matches 499; Conservative 0; Mismatches 3; Indels 2; Gaps 1;  
QY 1 AGATACCTCTTTGGAGAAACCATGTACCATGTGGTGACATTTGCTCTGGAAGGC 60  
DB 490 AGATACCTCTTTGGAGAAACCATGTACCATGTGGTGACATTTGCTCTGGAAGGC 549  
QY 61 TCTTTACGTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGAGGGCTGAGTTCA 120  
DB 550 TCTTTACGTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGAGGGCTGAGTTCA 609  
QY 121 GCGGAGGAGGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATGTTG 180  
DB 610 GCGGAGGAGGGCTGCCCGCTTCTGATTCATCAGCAAGCTATTCATCAATTAATGTTG 669  
QY 181 GGTGGAGGAACCGGGCGATGACGGTGAGAACTTTGTCAGTCAGTCCCAATGTTCT 240  
DB 670 GGTGGAGGAACCGGGCGATGACGGTGAGAACTTTGTCAGTCAGTCCCAATGTTCT 729  
QY 241 TGGCGTCATTTGCTTTCTCACCACCTTCCGCGAGGGCTGCTCAGGCAAAATATTCGATGG 300  
DB 730 TGGCGTCATTTGCTTTCTCACCACCTTCCGCGAGGGCTGCTCAGGCAAAATATTCGATGG 789  
QY 301 AATTCAGAGGGTTCGCCGCCAGCGGTAAGTCCCTGTCTGTGAGGGGATGTTGAACGT 360  
DB 790 AATTCAGAGGGTTCGCCGCCAGCGGTAAGTCCCTGTCTGTGAGGGGATGTTGAACGT 849  
QY 361 GAACAATATGATGCTCGGCAAGCTCAATGACATCGCCAGGAAACGCGGGCAGTCCT 420  
DB 850 GAACAATATGATGCTCGGCAAGCTCAATGACATCGCCAGGAAACGCGGGCAGTCCT 909  
QY 421 TGGCGCAGATGGCGCTTGCATGGTGTCTGCGCGCAAGGAGAGTACGGCGC--GGATACC 478



XX 25-SEP-2001.  
 XX 24-JUN-1998; 98US-0103840.  
 XX 24-JUN-1998; 98US-0103840.  
 XX (GENO-) INST GENOMIC RES.  
 XX Fleischmann RD, White OR, Fraser CM, Venter JC;  
 XX WPI; 2001-647261/74.  
 XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 XX 1551 and H37Rv differ  
 XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.  
 XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic  
 CC monitoring.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from USPTO  
 CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
 XX  
 SQ Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;  
 Query Match 5.1%; Score 121.2; DB 22; Length 4403765;  
 Best Local Similarity 49.5%; Pred. No. 1.7e-24;  
 Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;  
 QY 1433 CAACGTGACACTTGTCTCAATCATTTGATGAAGCGAGCTTGAAGCGCTCTCTTAGCC 1492  
 DB 2227172 CAGCTGGCGCATGTGCTGCCGTGTGCTGAACTGGGAGCTTGTGATCGCGCGGAGCGC 2227113  
 QY 1493 CTTTCCATTTCCCTCGCGGTGAGTCAAGCGCTTAAAGCTTCGAGCATCACTGGGT 1552  
 DB 2227112 CTACATGTCAACCGCTCGGTGTGCTGAGTCAAGTCTGTGAGCAGCAGGTTCGCG 2227053  
 QY 1553 CGAGTGTGTTATCGCGACCCCAACCGCCCAAGCAACCGAGCGGTGAGTCTTGTG 1612  
 DB 2227052 CAGGTGCTGTGTGCTCAGGGAAAGCCATGTGCGGCGACGACCGCAGGTATCCCGCTGTTG 2226993  
 QY 1613 CAACGACGCGGAAATGTGTCTCAAGCAGAACTAAAGCGCAACTATCTGG--AC 1670  
 DB 2226992 CGGTGGCGCGCAACAGCGTTGCTCGAGTCCGAGCGCTGCTGAATGGGTGGCAAC 2226933  
 QY 1671 GCCTTGTGAA--ATCCCGTTAAACCATCGCCATCAACGAGATTCGCTATCCATGG 1726  
 DB 2226932 CGCTGCTGAAACGACGCGGATCACATTTGCGTAAACCGCGATTCATCGGACATGG 2226873  
 QY 1727 TTTCTCCGTGTTTCAACGAGGTAGTCTTGTGGGTGGAGCAAGCTCAGCTGCGGTG 1786  
 DB 2226872 TTTTTCGCGCGTGTGACG-----GTCTCGCGCGAGCTCTGCTCGACGTTCCGATC 2226822  
 QY 1787 GAAGATGAAGCGCACATATTCTCTGCGCGTGGAGATGTTTATGAGCGGTAAAC 1846  
 DB 2226821 GAGACAGAGCAATTCGCGCGCTGTCTACGGAGAGGTGTGGCGATGGCGCGGTGACC 2226762  
 QY 1847 CGTGAAGCTAATCCGTCGGGGATGTGAAGTAGTAGAATCTGGAACCATCGCCACTTG 1906  
 DB 2226761 ACCGAGCGGAACCCGTCGCGGCTGCGCGGTGCACCCCGTGGGTGAATGCGTACCTTA 2226702  
 QY 1907 GCCATTGCAACCCCTCATTTGCGGGATGCCCTACATGTTGATGGAAACTAGATTGGGCT 1966

DB 2226701 CCAAGTGGCCAGCAGCGCCATTTCGTCCAGCGCCATATATCCAGCGGTTCACCTGCCCGCGCG 2226642  
 QY 1967 GCGATGCCCGTCTTACGCTTCCGATCCCAAGATGTGCTTCAAGACCGTGACCTGGACGCGG 2026  
 DB 2226641 GCGCTAAAGCTCCGTCACTGGCGTGGNATCGTGACGATGGGCTGCAGGACATGTTGGTG 2226582  
 QY 2027 CGC---GTCCGATGGTCTGTGGGGCGCAGCGCGTATCCATTGTCCCGTCGCGCGGAAGGT 2083  
 DB 2226581 CGTAAGGCGCTTTCGTGCGGCCATCACCAGACCGCAGCACTTGTGCCGACCAAGAGGCG 2226522  
 QY 2084 TTTTGTGAGGCAATTCGCGGAGCGCTTGTGGGACCTTCTCCGAAACCCCAAGCTGCT 2143  
 DB 2226521 TTCACCGCGCAGCGCGCGCGCTGGATGGGGCATGTTCCTCCGAGAACTGGCAGCA 2226462  
 QY 2144 CCCATGCTAAAAGCAGGAGAGTGTATCTCTCCGATGAGATACCCATTCGACACACCGCATG 2203  
 DB 2226461 TCTCGCTTGGCGATGGATCGTTCGTACGGGTCTGCGACATACACTCGACGTCCTCTC 2226402  
 QY 2204 TATTGGCAACGATGGCGCTTGAATCTAGATCTTAGCTAGACTCACAGACCGCGTCTGTT 2263  
 DB 2226401 TATTGGCAATCTCGAACTGGACATCCGATCATCGCGCAATTACCGACACGCTGAGG 2226342  
 QY 2264 GATGACGCAATCGAGGATTTGGCGCTTAGTTAC 2297  
 DB 2226341 CGCGCGCAAGCGGTCTGTATCCGGGCGCCAGCAAC 2226308  
 RESULT 14  
 AAI99682/C  
 ID AAI99682 standard; DNA; 4411529 BP.  
 XX AC AAI99682;  
 XX DT 15-JAN-2002 (first entry)  
 XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.  
 XX KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;  
 XX OS Mycobacterium tuberculosis; patient treatment; epidemic monitoring; ds.  
 XX PA (GENO-) INST GENOMIC RES.  
 XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;  
 XX WPI; 2001-647261/74.  
 XX Evaluating strain variation of Mycobacterium tuberculosis, comprises  
 PT determining the nucleotide sequence of the strain at positions in the  
 PT genome corresponding to positions where M. tuberculosis strains CDC  
 XX 1551 and H37Rv differ  
 XX Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.  
 XX The invention relates to evaluating strain variation within and between  
 CC different populations of the tuberculosis bacterial pathogen,  
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the  
 CC nucleotide sequence of the first strain at positions in the complete  
 CC sequence of the genome that correspond to positions that differ in the  
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and  
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of  
 CC M. tuberculosis and has valuable application in the fields of  
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic

CC monitoring.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from USPTO  
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.  
XX  
SQ Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;  
Query Match 5.1%; Score 121.2; DB 22; Length 4411529;  
Best Local Similarity 49.5%; Pred. No. 1.7e-24;  
Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;  
Qy 1433 CAATGGACATTTGCTCAATCATATGATGAAGCGCTTCGAGGCGCTCTTACCC 1492  
Db 2229873 CAGTGGCCCATATGGCTGCGTGGTGAATCTGGCGAGCTTCGATGCGCGCGGAGCGC 2229814  
Qy 1493 CTTTCCATTTCCCTCGCGCGTGAAGTCAAGCTTAAAGCTCTCGAGCATCAGTGGGT 1552  
Db 2229813 CTACATGTACCCCGTGGTGTGAGTCAAGCGATCAAGTCTGTGAGCAGCAGGTGCGG 2229754  
Qy 1553 CGAGTGTGTATCGGCACCCCAACCGGCCAAAGCAACCGAGCGGTGAAGTCTCTTGTG 1612  
Db 2229753 CAGTGTGCTGTGTGAGGAAAGCCATGTCGGCGGACGACCGCAGGTATCCCGCTGTG 2229694  
Qy 1613 CAAGCAGCGCGAAATGGTGTCTGTCGACGACAACTAAAGCGCAACTATCTG--AC 1670  
Db 2229693 CGTGTGCGCGCAACAGCGCTTGTCTGAGTCCGAGCGCTCGCTGAATGGGTGCGAAC 2229634  
Qy 1671 GCCTTGTCTGAA---ATCCCGTTAAACCATCGCCATCAACGAGATTCGCTATCCACATGG 1726  
Db 2229633 GCGTGTCTGTAACGACGCGGATACCATTCGCTGAACCGCATTCCTATGGCGCATGG 2229574  
Qy 1727 TTTCTCTCCGTTTCAACAGAGTAGTCTTTTGGGGTGGAGCAAGCTCACGCTCGCGCTG 1786  
Db 2229573 TTTTGGCGCGTGTGACG-----GTCTCGCGCAGCTCTGCTCGAGTTCGATC 2229523  
Qy 1787 GAAGATGAAGGACACATTAATCTTGTGCGCGTGGAGATGTTTAGAGCGGTAAAC 1846  
Db 2229522 GAGGACGAGACCATTCGCGCGCTGCTACGCGAGGTGTGGCGATGGCGCGGTGACC 2229463  
Qy 1847 CGTGAAGCTAATCCGTTGGGGATGTAAGTAGTAGTGAAGCTTGAACCATCGCCACTTG 1906  
Db 2229462 ACCGAGCGGAACCGGTGCGCGCTGCGCGGTGCAACCGTGGGTGAATGGCTACCTA 2229403  
Qy 1907 GCCATTCACACCCCTCATTCGCGGATGCCCTATCATGTTGTATGGGAACTAGATTGGCT 1966  
Db 2229402 CCAGTGGCCAGCAGGCGCATTCGTCCAGCGCATCTATCCGACGGTTCACGCGCGG 2229343  
Qy 1967 GCGATGCCGCTTACGCTTCGCTCCCAAGATGTGTTTCAAGACCGTGAACCTGAGCGG 2026  
Db 2229342 GCGGCTAAAGCTCCGTCACCTGGCGTGGAAATCGTGAACGATGGCTGAGGACATGTTGGTG 2229283  
Qy 2027 CGC---GTGATGCTCTGTGGGCGCAGCGCGTATCCATTTGTCGCTCGCGGAGGT 2083  
Db 2229282 CGTAAGGCTTTCGTCGCGCATCACAGACGACGCGCATTTGTTCCTCCGACGAGGGC 2229223  
Qy 2084 TTTGGTGAAGCAATTCGCGAGGCTTTGGTGGGGAATCTTCTCCGAAACCAAGCTGCT 2143  
Db 2229222 TTCACCGCGCAGCGCGCGCGGCTGGATGGGATGTTCCCGAGAGCTGGCAGCA 2229163  
Qy 2144 CCNATGTAAAGCAGAGAGATGATCTCTCTGATGAGATACCATTTGACACACCGATG 2203  
Db 2229162 TCTCCGCTTTCGCGATGGATGCTGTTGATGCGGTCTGCGACATACACCTCGACGCTCTC 2229103  
Qy 2204 TATTGGCAAGATGGCGCTGGATCTAGATCTAGTACATCAAGCGGCTGCTT 2263  
Db 2229102 TATTGGCAATGCTGGAATCTGACATCGCATCGCGCAATTACCGACAGGTGAGG 2229043  
Qy 2264 GATGCAAGATCGAGGATTTGCGGCTTAGTTAC 2297  
Db 2229042 GCGGCGGCAAGCGGTCTGTACCGGGCGCAGCAAC 2229009

AAAF71729/c  
ID AAF71729 standard; DNA; 1200 BP.  
XX  
AC AAF71729;  
XX  
DT 30-APR-2001 (first entry)  
XX  
Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:739.  
DE  
XX  
KW Corynebacterium glutamicum; carbon metabolism and energy production;  
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;  
KW fine chemical production; organic acid; proteinogenic amino acid;  
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
KW diagnosis; Corynebacterium diptheriae; evolutionary study; ds.  
XX  
OS Corynebacterium glutamicum.  
XX  
PN WO200100844-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-IB00943.  
XX  
PR 25-JUN-1999; 99US-0141031.  
PR 08-JUL-1999; 99DE-1031412.  
PR 08-JUL-1999; 99DE-1031413.  
PR 08-JUL-1999; 99DE-1031419.  
PR 08-JUL-1999; 99DE-1031420.  
PR 08-JUL-1999; 99DE-1031424.  
PR 08-JUL-1999; 99DE-1031428.  
PR 08-JUL-1999; 99DE-1031431.  
PR 08-JUL-1999; 99DE-1031433.  
PR 08-JUL-1999; 99DE-1031434.  
PR 08-JUL-1999; 99DE-1031510.  
PR 08-JUL-1999; 99DE-1031562.  
PR 08-JUL-1999; 99DE-1031634.  
PR 08-JUL-1999; 99DE-1032180.  
PR 09-JUL-1999; 99DE-1032227.  
PR 09-JUL-1999; 99DE-1032230.  
PR 09-JUL-1999; 99US-0143208.  
PR 14-JUL-1999; 99DE-1032924.  
PR 14-JUL-1999; 99DE-1032973.  
PR 14-JUL-1999; 99DE-1033005.  
PR 27-AUG-1999; 99DE-1040765.  
PR 31-AUG-1999; 99US-0151572.  
PR 03-SEP-1999; 99DE-1042076.  
PR 03-SEP-1999; 99DE-1042079.  
PR 03-SEP-1999; 99DE-1042086.  
PR 03-SEP-1999; 99DE-1042087.  
PR 03-SEP-1999; 99DE-1042088.  
PR 03-SEP-1999; 99DE-1042095.  
PR 03-SEP-1999; 99DE-1042123.  
PR 03-SEP-1999; 99DE-1042125.  
XX  
PA (BADI ) BASF AG.  
XX  
PI Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;  
XX  
XX WPI: 2001-061975/07.  
XX P-PSDB; AAB79612.  
XX  
PT New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
PT metabolism and oxidative phosphorylation protein for production or  
PT modulation of production of fine chemicals e.g. amino acids,  
PT carbohydrates or enzymes -  
XX  
PS Claim 3; Page 1192-1194; 1246pp; English.  
XX  
CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
CC metabolism and oxidative phosphorylation (SMP) proteins given in  
CC AAB79243 to AAB 79633 which are involved in carbon metabolism and

CC energy production. The C. glutamicum SMP gene can be used in vectors  
 CC (II) for expression in host cells and production or modulation of  
 CC production of fine chemicals, such as, an organic acid, a proteinogenic  
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,  
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty  
 CC acid, a diol, a carboxylate, an aromatic compound, a vitamin, a  
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins  
 CC (III) encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to  
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,  
 CC in evolutionary studies, in determining SMP protein regions required  
 CC for function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (i.e. ATP, NADPH).

XX

SQ Sequence 1200 BP; 272 A; 367 C; 294 G; 267 T; 0 other;

Query Match 4.3%; Score 101; DB 22; Length 1200;  
 Best Local Similarity 100.0%; Pred. No. 2e-20;  
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2274 TCGAGGAAATGGCGCCTTAGTTACTTCTGAAAGGTTTCAGGGTTTTCACCTTCTTGGCCC 2333

DB 1200 TCGAGGAAATGGCGCCTTAGTTACTTCTGAAAGGTTTCAGGGTTTTCACCTTCTTGGCCC 1141

QY 2334 GCAGGAATGGCGCAGGAGTAACACCTTCAGCAATGG 2374

DB 1140 GCAGGAATGGCGCAGGAGTAACACCTTCAGCAATGG 1100

Search completed: April 27, 2003, 02:57:01  
 Job time : 10751.4 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2003, 00:51:48 ; Search time 268.11 Seconds  
(without alignments)  
9634.930 Million cell updates/sec

Title: US-09-105-117K-3

Perfect score: 2374

Sequence: 1 agatactccttggagaaa.....gtaacaccttcagcaaatgg 2374

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2374	100.0	3309400	9	US-09-738-626-1
2	993	41.8	993	9	US-09-746-660A-47
3	870	36.6	870	9	US-09-738-626-3456
4	822	34.6	822	9	US-09-746-660A-51
5	708	29.8	708	9	US-09-738-626-3455
6	627	26.4	627	9	US-09-738-626-3454
7	121.2	5.1	909	9	US-09-894-844-12
8	93.2	3.9	988	10	US-09-974-300-77
9	82.6	3.5	894	10	US-09-815-242-6195
10	78	3.3	1077	9	US-09-738-626-1395
11	76.2	3.2	894	10	US-09-815-242-9644
12	73	3.1	903	10	US-09-815-242-4130
13	64	2.7	1277	9	US-09-971-536-15
14	57.6	2.4	762	10	US-09-815-242-7643
15	57	2.4	597	9	US-09-894-844-11
16	43.8	1.8	1973	10	US-09-864-761-3471
17	42.8	1.8	1879	9	US-10-114-170-142
18	40.4	1.7	1075	10	US-09-864-761-19241
19	40.4	1.7	1403	10	US-09-864-761-2513

Sequence 358, App  
Sequence 29, Appl  
Sequence 1034, Ap  
Sequence 358, App  
Sequence 20241, A  
Sequence 109, App  
Sequence 1672, Ap  
Sequence 12392, A  
Sequence 20, Appl  
Sequence 592, App  
Sequence 592, App  
Sequence 439, App  
Sequence 1, Appli  
Sequence 346, App  
Sequence 346, App  
Sequence 7775, Ap  
Sequence 34, Appl  
Sequence 34, Appl  
Sequence 34, Appl  
Sequence 1, Appli  
Sequence 5, Appli  
Sequence 5, Appli  
Sequence 398, App  
Sequence 188, App  
Sequence 398, App

1049 9 US-10-123-155-358  
399 10 US-09-861-893-29  
431 10 US-09-954-456-1034  
1049 9 US-10-123-155-358  
1635 10 US-09-864-761-20241  
1812 9 US-09-712-363-109  
948 9 US-09-738-626-1672  
424 10 US-09-960-352-12392  
867 9 US-10-123-155-20  
499 9 US-10-184-644-592  
499 9 US-10-184-634-592  
1068 9 US-09-738-626-439  
3309400 9 US-09-738-626-1  
671 9 US-10-184-644-346  
542 9 US-10-184-634-346  
996 10 US-09-815-242-7775  
4689 9 US-09-860-846-34  
4689 9 US-09-988-384B-34  
4689 10 US-09-861-289-34  
5828 9 US-09-430-029-1  
36778 9 US-09-860-846-5  
36778 10 US-09-861-289-5  
37948 9 US-09-988-384B-5  
542 9 US-10-184-644-398  
542 9 US-10-123-155-188  
542 9 US-10-184-634-398

#### ALIGNMENTS

##### RESULT 1

US-09-738-626-1  
; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

Query Match 100.0%; Score 2374; DB 9; Length 3309400;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGATACTCCTTTGGAGAAACCATGTACCATTCCTGACATTTGCGTCTGGAAGGC 60  
DB 1327595 AGATACTCCTTTGGAGAAACCATGTACCATTCCTGACATTTGCGTCTGGAAGGC 1327654  
QY 61 TCTTTACGTGGGTATTTCCTTCCTACGGTCCAGAGCTCACAGCGGCGGCTGAGTTTCAT 120



|||||  
Db 1327655 TCCTTACGTTGGGATTTCTTCTACGGTCCAGAGCTCACAGCGAGCGCGCTGAGTTCTAT 1327714  
QY 121 GCGGAGGAGGGCTGCGCGCTCTTGATTTATCATGACCCAGCTATTCCATCATTAATCGTTG 180  
Db 1327715 GCGGAGGAGGGCTGCGCGCTCTTGATTTATCATGACCCAGCTATTCCATCATTAATCGTTG 1327774  
QY 181 GTGGAGGAACCGGGCGATGACGGTGAGAACTTGTTGCACTGAGTCCAGTCCCAACAATGTCT 240  
Db 1327775 GTGGAGGAACCGGGCGATGACGGTGAGAACTTGTTGCACTGAGTCCAGTCCCAACAATGTCT 1327834  
QY 241 TGCGCTCATTTGCTTTCACCACTTGCGCAGGGGCTGCTCAOCCGACAAATATCTCGATGG 300  
Db 1327835 TGCGCTCATTTGCTTTCACCACTTGCGCAGGGGCTGCTCAOCCGACAAATATCTCGATGG 1327894  
QY 301 AATTCCAGAGGGTTCCCGCGCAGCCAGGCTAAGTCCCTGTCTGAGGGCATGTTGAACGT 360  
Db 1327895 AATTCCAGAGGGTTCCCGCGCAGCCAGGCTAAGTCCCTGTCTGAGGGCATGTTGAACGT 1327954  
QY 361 GAACAAATATTGATATGTCGCGAAGCTCAATGATCATCGCCAGGAACCGCGGCGATCACT 420  
Db 1327955 GAACAAATATTGATATGTCGCGAAGCTCAATGATCATCGCCAGGAACCGCGGCGATCACT 1328014  
QY 421 TGCGCAGATGGCGCTTGCAATGGGTGCTGCGCAGCAAGGAGAGTACGCGCGCGATACCGT 480  
Db 1328015 TGCGCAGATGGCGCTTGCAATGGGTGCTGCGCAGCAAGGAGAGTACGCGCGCGATACCGT 1328074  
QY 481 GACCAATTGATTTGATTTGGTGTCTGTCAGTTGAGCAGCTGGACAAACCGCTTGATTCAC 540  
Db 1328075 GACCAATTGATTTGATTTGGTGTCTGTCAGTTGAGCAGCTGGACAAACCGCTTGATTCAC 1328134  
QY 541 CAACAATTGAGTTTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACAGAGCG 600  
Db 1328135 CAACAATTGAGTTTCTGACGCGGAGTTGGAGGCGATCGATGAGATTTCCACAGAGCG 1328194  
QY 601 CGGCATCAACATTTGGGCGAAGGCCACCGATTCCAAACCCGCGAAGCACTAAACCCATCAA 660  
Db 1328195 CGGCATCAACATTTGGGCGAAGGCCACCGATTCCAAACCCGCGAAGCACTAAACCCATCAA 1328254  
QY 661 CATCAGTTGATGCGCAATGCGGTCTATCACAACTGCCACGACGAGCTTGATCCAGCGCCA 720  
Db 1328255 CATCAGTTGATGCGCAATGCGGTCTATCACAACTGCCACGACGAGCTTGATCCAGCGCCA 1328314  
QY 721 CACCTTGGGCTGACAGCGGGGTGACAAATGCTGCTGCGCGGAAACCCACAGCGGGAA 780  
Db 1328315 CACCTTGGGCTGACAGCGGGGTGACAAATGCTGCTGCGCGGAAACCCACAGCGGGAA 1328374  
QY 781 CCAGATCAGGCTTGCGCGAAGCGCGCAGCGGGAATAATCCACCGTCCGGTGTGCGCGTA 840  
Db 1328375 CCAGATCAGGCTTGCGCGAAGCGCGCAGCGGGAATAATCCACCGTCCGGTGTGCGCGTA 1328434  
QY 841 TTGCGCGCGACGCGCGCGATAACACAAACGGGTCCAAATACGCATTCGGGTTCAACCA 900  
Db 1328435 TTGCGCGCGACGCGCGCGATAACACAAACGGGTCCAAATACGCATTCGGGTTCAACCA 1328494  
QY 901 GGTGACAGCATGCGCATCAACATGCGCTTACCCAAACCGGTCTTATCGACGCTCAC 960  
Db 1328495 GGTGACAGCATGCGCATCAACATGCGCTTACCCAAACCGGTCTTATCGACGCTCAC 1328554  
QY 961 CTCACCCGCAACCGGTGCGCGTGCAGTGGCCACCGCGCAACCGCCCAAGCGGTGC 1020  
Db 1328555 CTCACCCGCAACCGGTGCGCGTGCAGTGGCCACCGCGCAACCGCCCAAGCGGTGC 1328614  
QY 1021 ATCGGCAACGGTGGTCTGTTCTCAATGATCTGCGGCTTCCACCTTGTGTCAT 1080  
Db 1328615 ATCGGCAACGGTGGTCTGTTCTCAATGATCTGCGGCTTCCACCTTGTGTCAT 1328674  
QY 1081 GGCGTCTTTCGCTGCGATGACGGAACCAATACAGATGAGCATGCGCCAGCGCAT 1140  
Db 1328675 GGCGTCTTTCGCTGCGATGACGGAACCAATACAGATGAGCATGCGCCAGCGCAT 1328734  
QY 1141 AATATCGAGCAGCATCGCGCGCATTTGGACAAAGATCAACGCCCAAGTTCGCGCGAT 1200  
|||||

Db 1328735 AATATCGAGCAGCATCGCGCGCATTTGGACAAAGATCAACGCCCAAGTTCGCGCGAT 1328794  
QY 1201 GAACAAAAAGACGTCAAAAATTAACACACGAGAGAAACCGCAATGAGTCTCTTCGCGCTT 1260  
Db 1328795 GAACAAAAAGACGTCAAAAATTAACACACGAGAGAAACCGCAATGAGTCTCTTCGCGCTT 1328854  
QY 1261 AATTCCTTGTAAATCAACACAGTACATTCGCGGTCCGATGGACAGTAAAGACTGCGCCC 1320  
Db 1328855 AATTCCTTGTAAATCAACACAGTACATTCGCGGTCCGATGGACAGTAAAGACTGCGCCC 1328914  
QY 1321 CAAAACGACAGCTGTAATCAAGATTTCCATGATCACCATCTGACCTATGGAGTACTTA 1380  
Db 1328915 CAAAACGACAGCTGTAATCAAGATTTCCATGATCACCATCTGACCTATGGAGTACTTA 1328974  
QY 1381 AGTAAATGATTTGTTCTTAAACATGGTTAAATATAGTCTTCATGAACCCCATTTCAAATGGA 1440  
Db 1328975 AGTAAATGATTTGTTCTTAAACATGGTTAAATATAGTCTTCATGAACCCCATTTCAAATGGA 1329034  
QY 1441 CACTTTGCTCTCAATCATTTGATGAAGGAGCTTCGAGGGGCTCTTAGCCCTTTCCAT 1500  
Db 1329035 CACTTTGCTCTCAATCATTTGATGAAGGAGCTTCGAGGGGCTCTTAGCCCTTTCCAT 1329094  
QY 1501 TTCCCTCTCGCGGTGAGTCAAGCGTTAAAGCTCTCGAGCATCACTGGGTGAGTGT 1560  
Db 1329095 TTCCCTCTCGCGGTGAGTCAAGCGTTAAAGCTCTCGAGCATCACTGGGTGAGTGT 1329154  
QY 1561 GGTATCGCGACCCCAACCGGCCAAAGAACCGAAGCGGTGAAGTCTTTGTGCAAGCAGC 1620  
Db 1329155 GGTATCGCGACCCCAACCGGCCAAAGAACCGAAGCGGTGAAGTCTTTGTGCAAGCAGC 1329214  
QY 1621 GCGGAAATGTTGTTGCTCTCAAGCAGAACTAAAGCGCAACTATCTGGAGCGCTTCTGA 1680  
Db 1329215 GCGGAAATGTTGTTGCTCTCAAGCAGAACTAAAGCGCAACTATCTGGAGCGCTTCTGA 1329274  
QY 1681 AATCCCGTTAAACCATCGCATCAACGAGATCGCTTATCAATGTTTCTCCCGTGT 1740  
Db 1329275 AATCCCGTTAAACCATCGCATCAACGAGATCGCTTATCAATGTTTCTCCCGTGT 1329334  
QY 1741 CAAACAGGTAGCTTCTTGGGGTGGAGCAACGCTCAACGCTTGGAAAGATGAAGCGCA 1800  
Db 1329335 CAAACAGGTAGCTTCTTGGGGTGGAGCAACGCTCAACGCTTGGAAAGATGAAGCGCA 1329394  
QY 1801 CACATTTCTTCTGCGCGGTGGAGATGTTTTAGAGCGGTAAACCGGTGAGCTAATCC 1860  
Db 1329395 CACATTTCTTCTGCGCGGTGGAGATGTTTTAGAGCGGTAAACCGGTGAGCTAATCC 1329454  
QY 1861 CGTGGCGGATGTGAAGTAGTAGAACTTGGAACCATGCGCACTTGCCCATTTGCAACCCC 1920  
Db 1329455 CGTGGCGGATGTGAAGTAGTAGAACTTGGAACCATGCGCACTTGCCCATTTGCAACCCC 1329514  
QY 1921 CTCATTTGCGGGATGCTTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTCTT 1980  
Db 1329515 CTCATTTGCGGGATGCTTACATGTTGATGGGAACTAGATTGGGCTGCGATGCCGCTCTT 1329574  
QY 1981 AGCTTTGGTCCCAAGATGCTTCAAGACCGTGAACCTGGAGCGGCGCTGATGGTCC 2040  
Db 1329575 AGCTTTGGTCCCAAGATGCTTCAAGACCGTGAACCTGGAGCGGCGCTGATGGTCC 1329634  
QY 2041 TGTGGGGCGCAGCGCGTATTCATTTCCGCTCGCGGGAAGTGTGTTGGTGAGGCAATTCG 2100  
Db 1329635 TGTGGGGCGCAGCGCGTATTCATTTCCGCTCGCGGGAAGTGTGTTGGTGAGGCAATTCG 1329694  
QY 2101 CGAGGCTTGTGTTGGGACTTCTTCCGAAACCCAAAGTGTCTCCATGCTTAAAGCAGG 2160  
Db 1329695 CGAGGCTTGTGTTGGGACTTCTTCCGAAACCCAAAGTGTCTCCATGCTTAAAGCAGG 1329754  
QY 2161 AGAGTGTCTCTCGATGAGTACCCATTGACACCGATGATTTGGTGCNACGATGGCG 2220  
Db 1329755 AGAGTGTCTCTCGATGAGTACCCATTGACACCGATGATTTGGTGCNACGATGGCG 1329814  
QY 2221 CTTGGAATCTAGATCTTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATCGAGGG 2280  
Db 1329815 CTTGGAATCTAGATCTTAGCTAGACTCACAGACGCGCTGTTGATGACGCAATCGAGGG 1329874

QY 2281 ATTGCGGCTTAGTTACTTCTGAAAGGTTTCAGGGTTTTCACCTTCTTCCGCGCAGGAA 2340  
Db 1329875 ATTGCGGCTTAGTTACTTCTGAAAGGTTTCAGGGTTTTCACCTTCTTCCGCGCAGGAA 1329934  
QY 2341 TTGGGCCAGCAGAGTAAACCTTCAGCAATGG 2374  
Db 1329935 TTGGGCCAGCAGAGTAAACCTTCAGCAATGG 1329968

## RESULT 2

US-09-746-660A-47  
; Sequence 47, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zeider, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; TITLE OF INVENTION: METABOLIC PATHWAY PROTEINS  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; CURRENT FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; PRIOR FILING DATE: 1999-07-08  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 47  
; LENGTH: 993  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (101)...(970)  
; OTHER INFORMATION: RXA01393  
US-09-746-660A-47

Query Match 41.8%; Score 993; DB 9; Length 993;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 993; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1321 CAAAGCAGCCTGTAATGAAGTTTCCATGATCACCATCGTGAACCTATGGAAGTACTTA 1380  
Db 1 CAAAGCAGCCTGTAATGAAGTTTCCATGATCACCATCGTGAACCTATGGAAGTACTTA 60  
QY 1381 AGTAAATGATGTTCTTAAACATGTTTAAATAGCTTCATGAACCCCACTCACTGA 1440  
Db 61 AGTAAATGATGTTCTTAAACATGTTTAAATAGCTTCATGAACCCCACTCACTGA 120  
QY 1441 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGCCTCCTTAGCCCTTTCCAT 1500  
Db 121 CACTTTGCTCTCAATCATGATGAAGCAGCTTCGAAGCGCCTCCTTAGCCCTTTCCAT 180  
QY 1501 TTCCCTCTGGCGGTGAGTCAGCGGTTAAAGCTTCGAGCATCAGTGGGTGAGTGT 1560  
Db 181 TTCCCTCTGGCGGTGAGTCAGCGGTTAAAGCTTCGAGCATCAGTGGGTGAGTGT 240

QY 1561 GGTATCGCGCACCCCAACCGGCCAAAGCAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 1620  
Db 241 GGTATCGCGCACCCCAACCGGCCAAAGCAACCGAAGCGGTGAAGTCTTGTGCAAGCAGC 300  
QY 1621 GCGGAAATGTTGTTGCTGCAAGCAGAAACTAAAGCGCAACTATCTGGAAGCCTTTGCTGA 1680  
Db 301 GCGGAAATGTTGTTGCTGCAAGCAGAAACTAAAGCGCAACTATCTGGAAGCCTTTGCTGA 360  
QY 1681 AATCCCGTTAACCATCGCCATCAACGCAGATTGCTATCCACATGTTTCTCCCGTGT 1740  
Db 361 AATCCCGTTAACCATCGCCATCAACGCAGATTGCTATCCACATGTTTCTCCCGTGT 420  
QY 1741 CAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTGGCTTGGAGATGAAGCGCA 1800  
Db 421 CAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTGGCTTGGAGATGAAGCGCA 480  
QY 1801 CACATTATCTTGTGCGCGGTGGAGATGTTTGGAGCGGTAAACCGGTGAAGCTAATCC 1860  
Db 481 CACATTATCTTGTGCGCGGTGGAGATGTTTGGAGCGGTAAACCGGTGAAGCTAATCC 540  
QY 1861 CGTGGCGGATCTGAAGTAGTAGAACTTGGAAACCATGGCCACTTGGCCATTGCAACCCC 1920  
Db 541 CGTGGCGGATCTGAAGTAGTAGAACTTGGAAACCATGGCCACTTGGCCATTGCAACCCC 600  
QY 1921 CTCATTGCGGGATGCTCATGCTTGTGAGGAAACTAGATTGGGCTGCGATGCCGCTCTT 1980  
Db 601 CTCATTGCGGGATGCTCATGCTTGTGAGGAAACTAGATTGGGCTGCGATGCCGCTCTT 660  
QY 1981 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTGAACCTGACCGCGCGCTCGATGGTCC 2040  
Db 661 ACGCTTCGGTCCCAAGATGCTTCAAGACCGTGAACCTGACCGCGCGCTCGATGGTCC 720  
QY 2041 TGTGGGCGCAGCGCGGTATCCATTGTCCTCGCGGGAAGGTTTTTGTGAGGCAATTGG 2100  
Db 721 TGTGGGCGCAGCGCGGTATCCATTGTCCTCGCGGGAAGGTTTTTGTGAGGCAATTGG 780  
QY 2101 CCGAGCGCTTGGTGGGGAATCTTCCGGAACCAAGCTCTCCCATGCTTAAAGCAGG 2160  
Db 781 CCGAGCGCTTGGTGGGGAATCTTCCGGAACCAAGCTCTCCCATGCTTAAAGCAGG 840  
QY 2161 AGAAGTGATCTCTCGATGATACCCATTGACACACCGATGATTGGCAACGATGGCG 2220  
Db 841 AGAAGTGATCTCTCGATGATACCCATTGACACACCGATGATTGGCAACGATGGCG 900  
QY 2221 CCTGGAATCTAGATCTTAGCTAGACTCAGACCGCGCTGTTGATGCAAGCTCGAGG 2280  
Db 901 CCTGGAATCTAGATCTTAGCTAGACTCAGACCGCGCTGTTGATGCAAGCTCGAGG 960  
QY 2281 ATTGCGGCTTAGTTACTTCTGAAAGGTTTCAAG 2313  
Db 961 ATTGCGGCTTAGTTACTTCTGAAAGGTTTCAAG 993

## RESULT 3

US-09-738-626-3456  
; Sequence 3456, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18

;; PRIOR APPLICATION NUMBER: JP 99/377484  
;; PRIOR FILING DATE: 1999-12-16  
;; PRIOR APPLICATION NUMBER: JP 00/159162  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: JP 00/280988  
;; PRIOR FILING DATE: 2000-08-03  
;; NUMBER OF SEQ ID NOS: 7059  
;; SOFTWARE: PatentIn ver. 3.0  
;; SEQ ID NO 3456  
;; LENGTH: 870  
;; TYPE: DNA  
;; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3456

Query Match 36.6%; Score 870; DB 9; Length 870;  
Best Local Similarity 100.0%; Pred. No. 2.4e-285;  
Matches 870; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1421 ATGAACCCCAATCACTGGACATTTGCTCTCAATCATGATGAAGCAGCTTCGAAGC 1480  
DB 1 ATGAACCCCAATCACTGGACATTTGCTCTCAATCATGATGAAGCAGCTTCGAAGC 60  
QY 1481 GCCTCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 1540  
DB 61 GCCTCTTAGCCCTTTCCATTTCCCTCGCGGTGAGTCAGCGCTTAAAGCTCTCGAG 120  
QY 1541 CATCAGTGGGTGAGTGTGGTATCGCGCACCCCAACCGGCCAAAGCAACCGAAGCGGT 1600  
DB 121 CATCAGTGGGTGAGTGTGGTATCGCGCACCCCAACCGGCCAAAGCAACCGAAGCGGT 180  
QY 1601 GAAGTCTTTGCAAGCAGCGCGGAAATGGTGTGCTGCAACGACGAACTAAAGCGAA 1660  
DB 181 GAAGTCTTTGCAAGCAGCGCGGAAATGGTGTGCTGCAACGACGAACTAAAGCGAA 240  
QY 1661 CTATCTGAGCGCTTGTGAAATCCCGTTAAACCATGCCATCAACGAGATTCGCTATCC 1720  
DB 241 CTATCTGAGCGCTTGTGAAATCCCGTTAAACCATGCCATCAACGAGATTCGCTATCC 300  
QY 1721 ACATGTTTCTCCCGTTCCTCAAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTG 1780  
DB 301 ACATGTTTCTCCCGTTCCTCAAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTG 360  
QY 1781 CGCTTGAAGATGAAGCGCACATATCTCTGCTGGCGGTGAGATGTTTAGGAGCG 1840  
DB 361 CGCTTGAAGATGAAGCGCACATATCTCTGCTGGCGGTGAGATGTTTAGGAGCG 420  
QY 1841 GTAAACCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTGAACCTTGAACCATGCGC 1900  
DB 421 GTAAACCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTGAACCTTGAACCATGCGC 480  
QY 1901 CATTGGCCATTGCAACCCCTCATTTGCGGGATGCCCTACATGTTGATGGGAACTAGAT 1960  
DB 481 CATTGGCCATTGCAACCCCTCATTTGCGGGATGCCCTACATGTTGATGGGAACTAGAT 540  
QY 1961 TGGGTGCGATGCGCTCTTACGCTTCGTTCCCAAGATGTGTTCAAGACCGTGACCTG 2020  
DB 541 TGGGTGCGATGCGCTCTTACGCTTCGTTCCCAAGATGTGTTCAAGACCGTGACCTG 600  
QY 2021 GACGGGCGCTCGATGCTCTGTGGGCGCAGCGGCTATCCATTGTCCGTCGGCGGAA 2080  
DB 601 GACGGGCGCTCGATGCTCTGTGGGCGCAGCGGCTATCCATTGTCCGTCGGCGGAA 660  
QY 2081 GGTGTTGGTGAGCAATTCGCGGAGGCTTGTGGTGGGACTTCTTCCGAAACCCCAAGCT 2140  
DB 661 GGTGTTGGTGAGCAATTCGCGGAGGCTTGTGGTGGGACTTCTTCCGAAACCCCAAGCT 720  
QY 2141 GCTCCCAATGCTTAAAGCAGAGATGATCTCTCGATGATATACCATTTGACACCG 2200  
DB 721 GCTCCCAATGCTTAAAGCAGAGATGATCTCTCGATGATATACCATTTGACACCG 780  
QY 2201 ATGATTTGGCAACGATGCGCGCTTGAATCTAGATCTCTAGCTAGACTCACAGACGCGTC 2260  
DB 781 ATGATTTGGCAACGATGCGCGCTTGAATCTAGATCTCTAGCTAGACTCACAGACGCGTC 840

QY 2261 GTTGATGAGCAATCGAGGATTCGGGCT 2290  
DB 841 GTTGATGAGCAATCGAGGATTCGGGCT 870

## RESULT 4

US-09-746-660A-51/c  
; Sequence 51, Application US/09746660A  
; Publication No. US20030049804A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompeius, Markus  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schroder, Hartwig  
; APPLICANT: Zeider, Oskar  
; APPLICANT: Haberhauer, Gregor  
; APPLICANT: Kim, Jun-Won  
; APPLICANT: Lee, Heung-Schick  
; APPLICANT: Hwang, Byung-Joon  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING  
; FILE REFERENCE: BGI-121CP2  
; CURRENT APPLICATION NUMBER: US/09/746,660A  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 09/606740  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 09/603124  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: 60/142101  
; PRIOR FILING DATE: 1999-07-02  
; PRIOR APPLICATION NUMBER: 60/148613  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/187970  
; PRIOR FILING DATE: 2000-03-09  
; PRIOR APPLICATION NUMBER: DE 19931420.9  
; NUMBER OF SEQ ID NOS: 125  
; SOFTWARE: PatentIn Vers. 2.0  
; SEQ ID NO 51  
; LENGTH: 822  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
; NAME/KEY: CDS  
; LOCATION: (101)..(799)  
; OTHER INFORMATION: RXA01394  
US-09-746-660A-51

Query Match 34.6%; Score 822; DB 9; Length 822;  
Best Local Similarity 100.0%; Pred. No. 5.3e-269;  
Matches 822; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 629 GATTCAAAACCCGCGAAACTAACCCATCAATCAGTTGATGCCCAATCGGTGATC 688  
DB 822 GATTCAAAACCCGCGAAACTAACCCATCAATCAGTTGATGCCCAATCGGTGATC 763  
QY 689 ACAACTGCCACACGACGTTGATCCAGGCCACACCTTGGGGCTGGACAGCGGGTGAC 748  
DB 762 ACAACTGCCACACGACGTTGATCCAGGCCACACCTTGGGGCTGGACAGCGGGTGAC 703  
QY 749 AATGCTGTGCGCCGAAACCCACCGCGGAAACCCAGATCAGGCTTTCGCCGCAACCGCCA 808  
DB 702 AATGCTGTGCGCCGAAACCCACCGCGGAAACCCAGATCAGGCTTTCGCCGCAACCGCCA 643  
QY 809 GCGCGGAAATCAACGTCGCTGTCGCGTATGGCGCCGACGCGCGCATTAACACA 868  
DB 642 GCGCGGAAATCAACGTCGCTGTCGCGTATGGCGCCGACGCGCGCATTAACACA 583  
QY 869 AACGCGTCAAAATACGCAATTCGGGTTCAACACAGGTGACGATGCGCATCAACATGGGC 928  
DB 582 AACGCGTCAAAATACGCAATTCGGGTTCAACACAGGTGACGATGCGCATCAACATGGGC 523

```
Qy 929 TTTACCCAAACCCGCTGCTTATCGACGCTCACCTCCACCCGACCCGCTTGCCTGCTCA 988
Db 522 TTTACCCAAACCCGCTGCTTATCGAGGCTCACCTCCACCCGACCCGCTTGCCTGCTCA 463
Qy 989 GTGGCCACCGCGCAACCGCCCAAGGCGTGTCTCGGCGACGCTGCTTCTGTTCTTCA 1048
Db 462 GTGGCCACCGCGCAACCGCCCAAGGCGTGTCTCGGCGACGCTGCTTCTGTTCTTCA 403
Qy 1049 ATGATCTGTGGCGCTTCCACCTGCTTGTGCTATGCGCTCTTTCGCTGCGCATGACGCGCAAC 1108
Db 402 ATGATCTGTGGCGCTTCCACCTGCTTGTGCTATGCGCTCTTTCGCTGCGCATGACGCGCAAC 343
Qy 1109 CATACAGGTAAGCGATGCGCACCCAGCGCATATATCGAGCAGATCGGCGCGCATTTG 1168
Db 342 CATACAGGTAAGCGATGCGCACCCAGCGCATATATCGAGCAGATCGGCGCGCATTTG 283
Qy 1169 GACAAAAGATCAACCGCCCAAGGCGCGGCGATGAACAAAAGAGCGTCAGAAATTAACAC 1228
Db 282 GACAAAAGATCAACCGCCCAAGGCGCGGCGATGAACAAAAGAGCGTCAGAAATTAACAC 223
Qy 1229 ACAGAAAGAACCGCAATGAGTCTTTCGCGCTTAATTCCTGTTTAAATCACCAGTACATTC 1288
Db 222 ACAGAAAGAACCGCAATGAGTCTTTCGCGCTTAATTCCTGTTTAAATCACCAGTACATTC 163
Qy 1289 TCGCGTCCGATGGACAGTAAAGACTGCGCCCAAGAGCAGACCTGTAATGAAGATTTCC 1348
Db 162 TCGCGTCCGATGGACAGTAAAGACTGCGCCCAAGAGCAGACCTGTAATGAAGATTTCC 103
Qy 1349 ATGATACCATCGTGACCTATGGAAGTACTTAAGTAAATGATTTGTTCTTAACATGGTT 1408
Db 102 ATGATACCATCGTGACCTATGGAAGTACTTAAGTAAATGATTTGTTCTTAACATGGTT 43
Qy 1409 TAATATAGTCTCATGAACCCCAATCACTGGACACTTTGCTC 1450
Db 42 TAATATAGTCTCATGAACCCCAATCACTGGACACTTTGCTC 1
```

## RESULT 5

US-09-738-626-3455/c  
; Sequence 3455, Application US/09738626  
; Publication No. US20020197605A1

## GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKOHO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3455  
; LENGTH: 708  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum

US-09-738-626-3455

Query Match 29.8%; Score 708; DB 9; Length 708;  
Best Local Similarity 100.0%; Pred. No. 3.4e-230;

```
Matches 708; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 652 ACCATCAACATCAGTTTGTATGCGCAATGCGGTGCATCAAACTGCGCACGACGTTGAT 711
Db 708 ACCATCAACATCAGTTTGTATGCGCAATGCGGTGCATCAAACTGCGCACGACGTTGAT 649
Qy 712 CAGCGCCACACCTTGGGGCTGGACAGCGGGCGTGA CAATGCTGCTGCGCCGAAACCCAC 771
Db* 648 CAGCGCCACACCTTGGGGCTGGACAGCGGGCGTGA CAATGCTGCTGCGCCGAAACCCAC 589
Qy 772 CAGCGGAACAGATCAGGCTTGC CGGAAACGCGCCAGCGGCGGAAATCAACGCTCCGTT 831
Db 588 CAGCGGAACAGATCAGGCTTGC CGGAAACGCGCCAGCGGCGGAAATCAACGCTCCGTT 529
Qy 832 GTCGCGTATGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 891
Db 528 GTCGCGTATGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 469
Qy 892 GTTCAACACAGGTCAGCAGCATGCGCATCAACATGGGCTTTACCCAAACCGCTGCTTATC 951
Db 468 GTTCAACACAGGTCAGCAGCATGCGCATCAACATGGGCTTTACCCAAACCGCTGCTTATC 409
Qy 952 GAGCTCACTCCACCGCGCGGCTTGC CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1011
Db 408 GAGCTCACTCCACCGCGCGGCTTGC CGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 349
Qy 1012 AGCGCGTATGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1071
Db 348 AGCGCGTATGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 289
Qy 1072 GTTGTGTCATGGGCTCTTTCGCTGCCATGCGCAACCAATTAACAGGTAAGCGATGCCACC 1131
Db 288 GTTGTGTCATGGGCTCTTTCGCTGCCATGCGCAACCAATTAACAGGTAAGCGATGCCACC 229
Qy 1132 CMAGCGCAATATCGAGCAGCATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1191
Db 228 CMAGCGCAATATCGAGCAGCATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 169
Qy 1192 GCGGCGGATGAACAAAGACGCTCAGAAATTAACACAGAGNAGNACCGCAATGATGTC 1251
Db 168 GCGGCGGATGAACAAAGACGCTCAGAAATTAACACAGAGNAGNACCGCAATGATGTC 109
Qy 1252 TTCGCGCTTAAATTCCTTGTGTTAATCACCAGTACATTCCTCGGTCCGATGCGAGTAAAG 1311
Db 108 TTCGCGCTTAAATTCCTTGTGTTAATCACCAGTACATTCCTCGGTCCGATGCGAGTAAAG 49
Qy* 1312 ACTGGCCCCCAAGCAGACCTGTAATGAAGATTTCCATGATCACCAT 1359
Db 48 ACTGGCCCCCAAGCAGACCTGTAATGAAGATTTCCATGATCACCAT 1
```

## RESULT 6

US-09-738-626-3454  
; Sequence 3454, Application US/09738626  
; Publication No. US20020197605A1

## GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKOHO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738, 626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162

; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 3454  
; LENGTH: 627  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3454

Query Match 26.4%; Score 627; DB 9; Length 627;  
Best Local Similarity 100.0%; Pred. No. 1.2e-202;  
Matches 627; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 23 ATGTACGCAATTTGGGTGACATTTGGTCTCTGGAAGGCTCTTTACGTGGGTATTTCTTCC 82  
DB 1 ATGTACGCAATTTGGGTGACATTTGGTCTCTGGAAGGCTCTTTACGTGGGTATTTCTTCC 60  
QY 83 TAGGTTCCAGAGCTCAGCGGAGGCGGCTGAGTTCTATGCGGAGGAGGCGCTCCCGCTT 142  
DB 61 TAGGTTCCAGAGCTCAGCGGAGGCGGCTGAGTTCTATGCGGAGGAGGCGCTCCCGCTT 120  
QY 143 CTGATTCATCAGCAAGCTATTCATCAATTAATGTTGGGTGAGGAACCGGCGATGAC 202  
DB 121 CTGATTCATCAGCAAGCTATTCATCAATTAATGTTGGGTGAGGAACCGGCGATGAC 180  
QY 203 GGTGAGAACTTTGTGAGTCTGAGTCCCAACAAATGCTTTGGCGTCAATTTCTCACCA 262  
DB 181 GGTGAGAACTTTGTGAGTCTGAGTCCCAACAAATGCTTTGGCGTCAATTTCTCACCA 240  
QY 263 CTTCGCGAGGCGCTCTCAGGCAAAATATCTGATGGAAATTCAGAGGTTCCCGGCGC 322  
DB 241 CTTCGCGAGGCGCTCTCAGGCAAAATATCTGATGGAAATTCAGAGGTTCCCGGCGC 300  
QY 323 AGCCAGGTAAGTCCCTGCTGAGGCGATGTTGAACGTGAACAAATATGATATGTCGCG 382  
DB 301 AGCCAGGTAAGTCCCTGCTGAGGCGATGTTGAACGTGAACAAATATGATATGTCGCG 360  
QY 383 AAGTCAATGACATCGCCAGGAAACCGGCGAGTCACTTCGCGAGATGGCGTTGATGG 442  
DB 361 AAGTCAATGACATCGCCAGGAAACCGGCGAGTCACTTCGCGAGATGGCGTTGATGG 420  
QY 443 GTGTCGCGAGCAAGAGAGTACGCGCGGATACCGTACCGAGTGAATGATGGTGTCT 502  
DB 421 GTGTCGCGAGCAAGAGAGTACGCGCGGATACCGTACCGAGTGAATGATGGTGTCT 480  
QY 503 TCGTCAGTTGAGCAGCTGGAACACAGCTTTGATTCACCAACTTGGAGTTTCTGAC 562  
DB 481 TCGTCAGTTGAGCAGCTGGAACACAGCTTTGATTCACCAACTTGGAGTTTCTGAC 540  
QY 563 GCCGAGTTGGAGGCGATCGATGAGATTTCCACGACCGCGCATCAACATTTGGGCGAAG 622  
DB 541 GCCGAGTTGGAGGCGATCGATGAGATTTCCACGACCGCGCATCAACATTTGGGCGAAG 600  
QY 623 GCCACCAATTCAAAACCCCGGAAAAC 649  
DB 601 GCCACCAATTCAAAACCCCGGAAAAC 627

## RESULT 7

US-09-894-844-12  
; Sequence 12, Application US/09894844  
; Patent No. US20020176873A1  
; GENERAL INFORMATION:  
; APPLICANT: Behr, Marcel  
; APPLICANT: Small, Peter  
; APPLICANT: Schoolnik, Gary A.  
; APPLICANT: Wilson, Michael A.  
; TITLE OF INVENTION: Molecular Differences Between Species of  
; TITLE OF INVENTION: the M. Tuberculosis Complex  
; FILE REFERENCE: STAN102CON  
; CURRENT APPLICATION NUMBER: US/09/894,844

; CURRENT FILING DATE: 2001-06-27  
; PRIOR APPLICATION NUMBER: 09/318,191  
; PRIOR FILING DATE: 1999-05-25  
; PRIOR APPLICATION NUMBER: 60/097,936  
; PRIOR FILING DATE: 1998-08-25  
; NUMBER OF SEQ ID NOS: 137  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 909  
; TYPE: DNA  
; ORGANISM: Mycobacteria tuberculosis  
US-09-894-844-12

Query Match 5.1%; Score 121.2; DB 9; Length 909;  
Best Local Similarity 49.5%; Pred. No. 3.8e-30;  
Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;  
QY 1433 CAACTGGACACATTTCTCTCAATCATTTGATGAAGCAGCTTCGAAGGCGCTCTTAGCC 1492  
DB 28 CAGCTGGCGCGCATTTGGTCCGTCGGAACCTGGGCGAGCTTCGATCGGCGCGGAGCGC 87  
QY 1493 CTTTCCATTTTCCCTCGGCGGTGAGTCAGCGCTTTAAAGCTCTCGAGCATCACGTGGGT 1552  
DB 88 CTACATGTCAACCCCGTGGCTGTCACTCAGCGCATCAAGTCTGTGGAGCAGCGGTGCGC 147  
QY 1553 CGAGTGTGGTATCGGCAACCCCAACCGGCCAAAGCAACGGAAGCGGCTGAAGTCTTGTG 1612  
DB 148 CAGGTGCTGTGTGTCAGGGAAAGCCATGTCGGGCGGACGACCGCAGGTATCCCGTGTG 207  
QY 1613 CAAGCAGCGCGAAATGTTGTCGCAAGCAGAAACTTAAAGCGCAACTATCTCG--AC 1670  
DB 208 CGTGGCGCGCGCAACAGCGTTGCTCGAGTCCGAGGCGCTCGCTGAATGGTGGCAAC 267  
QY 1671 GCCTTGCTGAA----ATCCCGTTAAACCATTCGCCATCAACGCAAGATTCGCTATCCACATGG 1726  
DB 268 GCGTGGCTGAAACGCGCGGATCACTTTCGGTAAACGCGGATTCATCGCGCATGG 327  
QY 1727 TTTCTCCCGTGTTCACAGAGTAGCTTCTTGGGGTGGAGCAACGCTACGCTCGGCTTG 1786  
DB 328 TTTTCGCGCGTGTTCGACG-----GTCTCGGCGACGCTCTGCTCGACGTTCCGATC 378  
QY 1787 GAAGTGAAGCGCACACATTTATCTTCTGTCGCGCGTGGAGATGTTTATAGAGCGGTAAAC 1846  
DB 379 GAGGACAGGACCATTTCCGCGCGCTGCTACGCGAGGGTGTGGCGATGGCGCGTGACC 438  
QY 1847 CGTGAAGCTTAATCCCGTGGCGGATGTGAAGTAGTAGAATTTTGAACCACTTGCCCACTTG 1906  
DB 439 ACCGAGCGGAACCCCGTGGCGGCTGCGGCTGCGCGCTGGTGAATGCGCTACCTA 498  
QY 1907 GCCATTGCAACCCCTCATTTGGGGGATGCTACATGGTTGATGGGAACTAGATTGGGCT 1966  
DB 499 CCAGTGGCGCAGCAGGCGCATTTCTGTCAGCGCCATCTATCCGACGGGTTCACTGCCCGCG 558  
QY 1967 GCGATGCCCGTCTTACGCTTCGCTCCCAAGATGTCTTCAAGACCGTGACCTGGACGG 2026  
DB 559 GCGGCTAAAGCTCCGTCATCTGGGTGGAATCTGACGATGGGCTGCGAGCATTTGGTG 618  
QY 2027 CGC---GTCGATGGTCTCTGTGGGCGCAGCGCGTATTCATTTGTCCCGCGGAAGGT 2083  
DB 619 CGTAAGGCTTTCTGTCGCGCCATCAACAGACGACGACATTTTGTCCCGACCAAGAGGCG 678  
QY 2084 TTTGTGGAGCAATTCGCGAGGCTTGGTTGGGAGCTTCTTCCCGAAACCCAGCTGCT 2143  
DB 679 TTTACCGCGCGCAGCGCGCGCGGCTGGGATGGGCGATTTCCCGGAGAGTGGGAGCA 738  
QY 2144 CCGATGCTAAAGCAGGAGAGTGTCTCTCGATGATATCCCATTCGACACACCGATG 2203  
DB 739 TCTCCGCTTGGCGATGGATCGTTCTGTACGGGTCTGGACATACACCTCGAGCTCTCTC 798  
QY 2204 TATTTGCAACGATGGCGCTCGGAATCTAGATCTCTAGCTAGACTCAGACGCGCGTGT 2263  
DB 799 TATTGCAATGCTGGAACCTGGACAGTCCGATCATCGCGCAATTAACGACACGCGTGAG 858

Qy 2264 GATGAGCAATCGAGGGATTGCGCTTAGTTAC 2297  
Db 859 GCGGCGCAAGCGTCTGTACCGGGCCAGCAAC 892

## RESULT 8

US-09-974-300-77  
; Sequence 77, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Clausen, Randy M.  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene  
; FILE OF INVENTION: Expression  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974.300  
; CURRENT FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 77  
; LENGTH: 988  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-77

Query Match 3.9%; Score 93.2; DB 10; Length 988;  
Best Local Similarity 56.0%; Pred. No. 1.4e-20;  
Matches 251; Conservative 0; Mismatches 173; Indels 24; Gaps 3;  
Qy 3 ATACTCTCTTGGAGAACCATGATGCGATTCGCTGACATGTTGGCTCTGGAAGCTC 62  
Db 428 ATACCGCTGGAAGACGATGGAGCGCTCGCCCATCGTCAGGCAAGGAAAGCTT 487  
Qy 63 TTTACGTGGGTATTTCTTCTACGGTTCAGAGCTTCAGAGCGGCGCTGAGTTCAATGG 122  
Db 488 TATATGTCGGTTTGTGCAACTACAGTGTGAACAGACCGGAAAGCGGCTGATTTTC 547  
Qy 123 CGGAGGAGGCTGCGGCTTCTGATTCATCAGCAAGCTATTCATTAATGTTGGG 182  
Db 548 AAGAATACGGGATTCGCTCTCTATCCATCAGCCCAATATTCATGTTCCACAG----- 602  
Qy 183 TGGAGGAACCGGCGATGACGGTGAGAACTTTGTCAGTCACTGCCCAACAAATGCTTGG 242  
Db 603 -----AGAACCGGAAAGGCGCTGCAAGTGTGCGAG-----AAAAAGGAGCGG 649  
Qy 243 GCGTCATTGCTTTCTACCACTTGGCAGGCGCTGCTCAGGACAAATATCTCGATGAA 302  
Db 650 GGGCGATCGGTTTACACCTCTTCCCAAGGGCTTTTAAACAGAAATATTTGACAGGA 709  
Qy 303 TTCAGAGGTTCCGCGCGCAGCCAGG-----GTAAAGTCCCTGCTGAGGGCATGTTGA 356  
Db 710 TTCCCGCGGGTCAAGAGCCGCTGATGCAACAGTCCGTTTTTCAGGCGCGAGGCATCA 769  
Qy 357 ACGTGAACATATTGATATGTTGTCGCAAGCTCAATGACATGCCCCAGAGCGGGCAGT 416  
Db 770 CAGAAGAGTCTCGGCAAGTTCGCAAGCTGATCAAGTGGTGGCCGAAACGCGGGCAGT 829  
Qy 417 CACTTGGCGAGATGGCGCTTGCATGGGT 444  
Db 830 CTCCTCCGCAATGCGATAGCTGGCT 857

## RESULT 9

US-09-815-242-6195  
; Sequence 6195, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Olesen, Kari L.

; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815.242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6195  
; LENGTH: 894  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(894)  
US-09-815-242-6195

Query Match 3.5%; Score 82.6; DB 10; Length 894;  
Best Local Similarity 48.8%; Pred. No. 5.3e-17;  
Matches 223; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

Qy 1455 TCATTGATGAAGGACGCTTCGAAGGCGCTCTTACGCCCTTTTCCATTTCCCTCGGCG 1514  
Db 44 TGATACGTGAACGAGGATTTGAGCGCGCGCACAAAGCTGTGCATTACACAATCAGCG 103  
Qy 1515 TGAGTCAGCGCGTTAAAGCTCTCGACATCACTGGGTGAGTTGGTATCGGCACCC 1574  
Db 104 TCTCAGAGCGCATTAAGCAACTGGAATATGTTGGGCGAGCGCTGTTGGTGTGCTACCG 163  
Qy 1575 AACCGGCCAAAGCAACCGAAGCGGGTGAAGTCTTGTGCAAGCAGCGCGGAAATGCTGT 1634  
Db 164 TACCGCGCGCGCGAGCGGAACAGGGGCAAAACTGCTGGCACTGCTGCCAGGTGGAGT 223  
Qy 1635 TGCTGCAAGCAAAACTAAAGCGCAACTATCTGGAGCGCTTCTGAAATCCCGTTAAACA 1694  
Db 224 TGCTGGAAGAGAGTGGCTGGCGATGAAACAAACCGTTTCGACTCCGCTGCTGCTTAC 283  
Qy 1695 TCGCATCAACGAGATTCGCTATCCATATGTTTCTCCCGTGTTCACAGGAGTAGCTT 1754  
Db 284 TGGCGGTCAACGCCGACAGTCTGGCGACGTGTTGTTCTCTCACTGGCTCTCTGTTGG 343  
Qy 1755 CTTGGGGTGGAGCAACGCTCACGCTCGCTTGGAGATGAAGCGCACACATATCTCTTGC 1814  
Db 344 CTGATTGCGCTATCCGCTCAACTTGCAGGTAGAAGATGAACCCGCACTCAGGAACGTC 403  
Qy 1815 TGGCGGTGGAGATGTTTGGAGCGGTAAACCCGTAAGCTAATCCCGTGGCGGATGTG 1874  
Db 404 TCGCGCGCGCGAAGTGGTGGCGGGTGAGTATTCAACATCAGGGCTGCCGAGTTGTC 463  
Qy 1875 AAGTAGTAAGACTTGAACCATCGGCCACTTGGCCAT 1911  
Db 464 TTGTCGATAAACTTGGTGGCTCGACTATCTGTGCT 500

## RESULT 10

US-09-738-626-1395/c  
; Sequence 1395, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAKO  
; APPLICANT: SENOH, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; APPLICANT: OZAKI, AKIO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; CURRENT FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: Patent in ver. 3.0  
; SEQ ID NO 1395  
; LENGTH: 1077  
; TYPE: DNA  
; ORGANISM: *Corynebacterium glutamicum*  
US-09-738-626-1395

Query Match 3.3%; Score 78; DB 9; Length 1077;  
Best Local Similarity 100.0%; Pred. No. 2.2e-15; Indels 0; Gaps 0;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2297 CTTCTGAAAGGTTACAGGTTTTCACCTTCTGCGCGCAGGAATGGGCGCAGCAGGT 2356  
DB 1077 CTTCTGAAAGGTTACAGGTTTTCACCTTCTGCGCGCAGGAATGGGCGCAGCAGGT 1018  
QY 2357 AACACCTTCAGCAATGG 2374  
DB 1017 AACACCTTCAGCAATGG 1000

## RESULT 11

US-09-815-242-9644  
; Sequence 9644, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq For Windows Version 4.0  
; SEQ ID NO 9644  
; LENGTH: 894  
; TYPE: DNA  
; ORGANISM: *Salmonella typhi*  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(894)  
US-09-815-242-9644

Query Match 3.2%; Score 76.2; DB 10; Length 894;  
Best Local Similarity 44.8%; Pred. No. 8e-15;

Matches 340; Conservative 0; Mismatches 413; Indels 6; Gaps 1;  
QY 1423 GAAACCCATTCAACTGGACACTTTGCTCTCAATCATTTGATGAAGGCGAGCTTCGAAGCGC 1482  
DB 12 GGACTACAGAACACTACAGGCGCTGATCGCTTATACGTGAACGAGGATTCGAGCGCGC 71  
QY 1483 CTCTTAGCCCTTTCCATTTCCCTCGCGGCTGAGTCAGCGGTTAAAGCTCTCGAGCA 1542  
DB 72 CGCCCAAGAGCTGTGCATCAGCAATCCGCCGTCTCACAGCGCATAAAAACAGCTTGAGAA 131  
QY 1543 TCACGTGGGTCGAGTGTGTTGTTATCGCGCACCAACCGCGGCAAGCAACGAAAGCGGTGA 1602  
DB 132 TATGTTGGAGCAACCTTTGTTGGTGCTACCGTTCTCTCGCCCCACGAGCAAGGGGA 191  
QY 1603 AGTCCTTGTGCAAGCAGCGCGGAAATGTTGTGTGCAAGCAGAACTTAAAGCGCACT 1662  
DB 192 AAAATTAATGCTCTGCTGCGTCAGCTTGAACCTGTGGAAGAGAGTGCTGGCGATGA 251  
QY 1663 ATCTGAGCGCTTGTGAATCCGTTAACCATGCGCATCAAGCGCAGATTCGTTATCCAC 1722  
DB 252 ACAACCGGCTCTACGCGCTACTGCTGCTGCGGCGGTTAAACGCGCAGACTCTGCGCAC 311  
QY 1723 ATGGTTTCTCCCGTGTCAACGAGGTAGCTTTCTTGGGGTGGAGCAACGCTCACGCTGG 1782  
DB 312 CTGGCTGCTTCCGGCGCTGGCGCCAGTACTGGGGAATCCCTATCCGTTCAATTACA 371  
QY 1783 CTTGGAAGATGAAGCGCACACATTAATCTTGTGCGGGCTGGAGATGTTTTAGGAGCGGT 1842  
DB 372 GGTGGAAGATGAACGCGCACAGACGCTCTGCGTCCGCGCGAGGTGGTGGCGCGGT 431  
QY 1843 AACCCGTGAAGCTAATCCCGTGGCGGATGTGAAGTAGTAGTAACCTGGAACTATGGCCA 1902  
DB 432 GAGTATTCAGCATCAGGCGCTGCCAGTTGCTGTGTAGACAACTGGGCGCGCTCGACTA 491  
QY 1903 CTTGGCCATTGCAACCCCTCATTTGGGATGCTTACATGTTGATGGGAAACTAGATTG 1962  
DB 492 TCTGTTTGTGCGGTCAAGCCGTTTCCCGAGCGCTATTTTCCCAATGGGTACCCGGTC 551  
QY 1963 GGTGCGAT-----GCCCGTCTTAGCTTCGTTCCGTCCTCAAGATGTGCTTCAAGACCGTA 2016  
DB 552 GTCGCTGTGAAAGCGCGCGCGTTCGCGTTGACACCATCTGGATGATATGCATCAGGCTTT 611  
QY 2017 CTTGAGCGGCGGCTGATGTTGCTGTGGGGCGCAGGCGGCTATCCATTTGTCCTCGGC 2076  
DB 612 TTTGCAACAGAAATTTGATCTACCGCGCGCAGCGCTGCCCATATCGTTAAATTCATC 671  
QY 2077 GGAAGGTTTTGTGAGGCAATTCGCGAGCGCTTGTGTTGGGACTTCTTCCCGAAACCCA 2136  
DB 672 CGAAGGTTTTGTACAGCTTGGCGGTGAGGCGCAGGCGCTGCTGCATGATCCCTCACCTGA 731  
QY 2137 AGCTGCTCCCATGCTTAAAGCAGGAGAGTATCTCTCT 2175  
DB 732 GATTGAGAAAGAGCTGGAAGCGCGCGAGTTGATCAACTT 770



## RESULT 12

US-09-815-242-4130  
; Sequence 4130, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4130  
; LENGTH: 903  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-4130

Query Match 3.1%; Score 73; DB 10; Length 903;  
Best Local Similarity 46.0%; Pred. No. 9.9e-14;  
Matches 413; Conservative 0; Mismatches 460; Indels 24; Gaps 4;  
QY 1418 TTCAATGAACCCCACTTCACTGACACTTGTCTCTCAATCATTTGATGAAGCAGCTTCGAA 1477  
DB 7 TTTGTCACTACAAAGTTGCTCGCCGCGTGGCGCGTGGTGGAGCGGGGTTTCGAG 66  
QY 1478 GGCGCTCTCTTAGCCCTTTCCATTTCCCTCGCGCGTGGTGAATCAAGCGCTTAAAGCTCTC 1537  
DB 67 CGCGCCGACACAGGCGCTGGGCTGTGCAATCGCGGTATCGCAGCGGATCAAGCTGCTC 126  
QY 1538 GAGCATCAGTGGTCAAGTTGGTATCGCGCACCAACCGCGCAAGCAACCGAAGCG 1597  
DB 127 GAGCGCGGGTTCGCGCAAGCGGCTCTGTGGCGGAGACGCGCGCATCCCGCATCTC 186  
QY 1598 GGTGAAGTCTTTGCAAGCAGCGCGCAAAATGTGTGTGTCGACAGCAAACTAAG-- 1655  
DB 187 GGGCGCCCTTGTCTCAACCATGTGACAGGTGGCGGTGTGAGGGCGCCTGCGACGCG 246  
QY 1656 -----CGCAACTATCTGACGCGCTTGTGTAATCCGTTAAACCATCGCATCAACGCA 1708  
DB 247 TGGTGCAGAACTTGGATGAAGGGGTGCCCCGGAACGCTGCGCATCGCTCTCAACGCC 306  
QY 1709 GATTCGTATCCATCGTTCTCCCGTGTTCACGAGTAGTCTTCTGGGGTGGAGCA 1768  
DB 307 GATAGCTGGCGCACTGTGGCGCGCGGTGGCGGATTTCTGCGCGCAGCGCGGTG 366  
QY 1769 ACGTCAAGTGGCGTGGAGATGAAGCGCACATATTCCTTCTGCGCGGTGGAGAT 1828  
DB 367 CTGCTGAGACCTTGTGGTGGAGGACACAGAGGTAGGGCTCAAGCGCATGCGCGGGCGAA 426  
QY 1829 GTTTTAGAGCGGTAAACCCCGTAGAGCTAAATCCCGTGGCGGGATGTGAAGTAGTAGAACTT 1888

DB 427 GTGCGCGCTGCTGCTGCGGTAGCGCGCGCGTGGCGCGCGGTTCGCTGCTGCTC 486  
QY 1889 GGAACCATGCGCCACTTGGCCATTTGCAACCCCTCAAT-----GCGGATGCTTACATG 1942  
DB 487 GGAGCCATGCTTATCGCGCGCTGGCCAGCCCGGATTTCATTGCCCGCACATTTCCCGCG 546  
QY 1943 GTTGATGGGAACACTAGATTGGGCTGCGATGCGCGTCTTACGCTTCCGTTCCCAAGATG 2002  
DB 547 GGGTCAAGCGCGCCCTTGGCGCGGTGCGCGCATCGTGTTCGCGCGCGGACGACCTG 606  
QY 2003 CTTCAAGACCGTGACCTGGACGCGCGCTCGATGTCTCTGTGGGCGCGAGGCGGTATCC 2062  
DB 607 CTGCAACACCGCTTCTCAAGGACCTCGCG-----TCGAGGGCGGTTTCATCCACAC 660  
QY 2063 ATTGTCCGTCGCGGGAAGTTTGTGAGGCAATTCGCGGAGCCCTTGGTTGGGAGCTT 2122  
DB 661 CTCGCGCGTCTCGGAAGTTTCTGTCGCGCTCACCGCGCGCGGTCTCGGCTGGGCGCTG 720  
QY 2123 CTTCCCGAAACCCCAAGCTGCTCCCATGTCTAAAGCAGGAGAGTGTATC---CTCCTCGAT 2179  
DB 721 GTGCCGAGCGCGAGTGAAGCGAGCTGGCGCGCGGAACTGGTCAAGCTGCTGCGCC 780  
QY 2180 GAGATACCATTTGACACACCGCATGTATTGGCAACCATGCGCTGGAATCTAGATCTTA 2239  
DB 781 GGCCAGGTCTACGACGTACCGTTGTACTGGCACTACTGCGCAACCGCGCGCACTGCTC 840  
QY 2240 GCTAGACTCACAGACGCGCTGTTGATGAGCAATCAGGAGTTGCGGCTTAGTTA 2296  
DB 841 GCGTCTGCTCACCGAGCACCTGCTGCGCGCGCGGCTTGGTGGGGTGCTCA 897  
RESULT 13  
US-09-971-536-15  
; Sequence 15, Application US/09971536  
; Patent No. US20020159976A1  
; GENERAL INFORMATION:  
; APPLICANT: Glenn, Matthew  
; APPLICANT: Havukkala, Ilkka  
; APPLICANT: Bloksberg, Leonard  
; APPLICANT: Lubbers, Mark  
; APPLICANT: Dekker, James  
; APPLICANT: Christensen, Anna  
; APPLICANT: Holland, Ross  
; APPLICANT: O'Toole, Paul  
; APPLICANT: Reid, Julian  
; APPLICANT: Coolbear, Timothy  
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods  
; FILE REFERENCE: 1043c2  
; CURRENT APPLICATION NUMBER: US/09/971,536  
; CURRENT FILING DATE: 2001-10-02  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238  
; PRIOR FILING DATE: 2000-08-08  
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160  
; PRIOR FILING DATE: 2001-08-08  
; NUMBER OF SEQ ID NOS: 83  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 15  
; LENGTH: 1277  
; TYPE: DNA  
; ORGANISM: Lactobacillus rhamnosus  
US-09-971-536-15

Query Match 2.7%; Score 64; DB 9; Length 1277;  
Best Local Similarity 47.9%; Pred. No. 1.5e-10;  
Matches 279; Conservative 0; Mismatches 270; Indels 33; Gaps 2;  
QY 9 CTTTGAAGAAACCATGTACGCAATTCGCGTACATTTGCGTCTCGAAAGGCTCTTTAG 68  
DB 608 CTTTGAAGAGAGCGTTAATGCGCTGGATCAACCGTGGTACCGTGAAGGCTGTGATA 667











```
QY 1493 CTTTCCATTTCCTCCCGGTGAGTCAGCGGTTAAAGCTCTCGAGCATCAGCTGGT 1552
Db 5313 CTACATGTCACCCCTCCGCTGTGAGTCAGCGCATCAAGTCGTTGGAGCAGAGCTGGC 5372
QY 1553 CGAGTGTGGTATCGCGCACCCCAACCGGCCAAAGCAACCGAAGCGGTGAAGTCTTGTG 1612
Db 5373 CAGTGTGTTGTCGAGGAAGCCATGTCGGGCGAGCAGCGAGTATCCCGCTGTTG 5432
QY 1613 CAAGCAGCGCGGAAATGTTGCTGCAAGCAGAACTAAAGCGCAACTATCTGG--AC 1670
Db 5433 CGGTTGGCGCGCAACAGAGGTTGCTCGAGTCCGAGCGCTCGTGAATGGGTGGCAAC 5492
QY 1671 GCCTTGTCTGAA---ATCCGCTTAACCATGCCATCAACGAGATTCGCTATCCATGG 1726
Db 5493 CGCTCGCTGAACCGCACCGCATCACCATTGCGGTAACCGCGATTCATGGGACATGG 5552
QY 1727 TTTCTCTCCGCTGTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACGCTCAGCGTGGCTTG 1786
Db 5553 TTTTCGGCGGTGTTGAGC-----GTCTCGCGCACTGCTCGAGCTTCGATC 5603
QY 1787 GAAGATGAAGGCGACATATATCTCTGCGGGTGGAGATGTTTGGAGCGGTAAAC 1846
Db 5604 GAGGACGAGGACCATTCGCGCGCTCTACGCGAGGCTGTGGAGTGGCGCGGTGACC 5663
QY 1847 CGTGAAGCTAATCCGTCGCGGATGTGAAGTAGTAGAATCTTGAACCATGCGCCACTTG 1906
Db 5664 ACCGAGCGGAACCGGTGCGCGCTGCGGGTGACCCGCTGGGTGAATGCGCTACCTA 5723
QY 1907 GCCATTGCAACCCCTCATTTGCGGGATGCTACATGTTGATGGGAACTAGATGGGCT 1966
Db 5724 CCAGTGGCCAGCAGGCAATTCGTCAGCGCATCTATCCGACGGTTCACTGCGCGCG 5783
QY 1967 GCGATGCCGCTTACGCTTGGTCCCAAGATGTGCTTCAAGACCGTGACCTGAGCGGG 2026
Db 5784 GCGGCTAAAGCTCGTCACTGCGGTGGAATCGTGACATGGGCTGAGGACATGTTGGTG 5843
QY 2027 CGC---GTCCGATGCTGTGGGCGCGAGCGGTATCCATTCTCCGTCGGCGGAGGT 2083
Db 5844 CGTAAGCCCTTGTTCGCGCATACACAGACCGACGCACTTTGTCGCGACCAAGAGGC 5903
QY 2084 TTTGGTGAGGCAATTCGCGAGGCTTGTGTTGGGAGCTTCTTCCGAAACCCAAAGCTGCT 2143
Db 5904 TTCACCGCGCAGCGCGCGCTGGATGGGCGATGTTCCCGAGAGCTGGCAGCA 5963
QY 2144 CCATGCTAAAGCAGAGAGATGATCTCTCGATAGATACCCATTGACACACCGATG 2203
Db 5964 TCTCCGCTTCCGATGATCGTTCTGACGGGTCTGGACATACACCTCGACGTCCTCTC 6023
QY 2204 TATTGGCAACGATGGCGCTGGAACTAGATCTCTAGCTAGACTCACAGACCGCTCGTT 2263
Db 6024 TATTGGCACTGCTGGAACCTGGACAGTCCGATCATCGCGCAATTAACGACCGGTGAGG 6083
QY 2264 GATGAGCAATCGAGGATTCGCGCTTAGTTAC 2297
Db 6084 GCGCGGCAAGCGCTGTGTACCGGGCGGCAAC 6117
```

## RESULT 2

```
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 5.1%; Score 121.2; DB 4; Length 4403765;
Best Local Similarity 49.5%; Pred. No. 7.1e-24;
Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;

QY 1433 CAACTGGACATTTGCTCTCAATCAATTGATGAAGCAGCTTCGAAGGCGCTCTCTTAGCC 1492
Db 2227172 CAGCTGGCGCATTTGGCTCGCTGCTGCAACTGGGAGCTTCGATGCGCGCGGAGCGC 2227113
QY 1493 CTTTCCATTTCCCTCCGCGGTGAGTCAGCGCTTAAAGCTCTGAGCATCAGTGGT 1552
Db 2227112 CTACATGTCACCCCGTGGCTGTGAGTCAGCGCATCAAGTCGTTGGAGCAGCAGTGGC 2227053
QY 1553 CGAGTGTGGTATCGGCACCCCAACCGGCAAGCAACCGAAGCGGTGAAGTCTCTTGTG 1612
Db 2227052 CAGTGTCTGGTGTGTCAGGGAAGCCATGTGGGCGAGCAGCGAGTATCCGCTGTG 2226993
QY 1613 CAAAGCAGCGCGGAAATGTTGTTGTCGAAGCAGAAACTAAAGCGCAACTATCTGG--AC 1670
Db 2226992 CGGTTGGCGCGCAACAGCGCTTCTGAGTCCGAGCGCTCGCTGAAATGGGTGGCAAC 2226933
QY 1671 GCTTGTCTGAA---ATCCGCTTAACCATCGCCATCAAGCAGATTCGATATCCACATGG 1726
Db 2226932 CGCTGCTGTAACCAACGACGCGATCACCATTGCGGTAAACGCGGATTCATGGCGACATGG 2226873
QY 1727 TTTCTCTCCGCTTCAACGAGGTAGCTTCTTGGGGTGGAGCAACCGCTCAGCGTCGCTTG 1786
Db 2226872 TTTTCGCGCGTGTTCGAG-----GTCTCGCGCAGCTCTCTCGAGTTCGATC 2226822
QY 1787 GAAGATGAAGCGCACACATTTATCTTGTGCGGCGTGGAGATGTTTTAGGAGCGGTAAAC 1846
Db 2226821 GAGGACCAAGGACCATTCGCGCGCTGCTACGCGAGGCTGTGGCGATGGCGCGGTGACC 2226762
QY 1847 CGTGAAGCTAATCCGTCGCGGATGTGAGTAGTAGTAAGTGGAACTGGAACCATGCGGCACCTG 1906
Db 2226761 ACCGAGCGGAACCGCGTGCCTGGGCTGCACCGCTGCGGTGAATTCGCTACCTA 2226702
QY 1907 GCATTGCAACCCCTCATTTGCGGATGCTTACATGTTGATGGGAACTAGATTTGGCT 1966
Db 2226701 CCAGTGGCCAGCAGGCAATTCGTCAGGCGCATCTATCCGACGGGTTCACTGCGCGCGC 2226642
QY 1967 GCGATGCCGCTTTACGCTTCGGTCCCAAGATGTGCTTCAAGACCGGTGACCTGGACCGG 2026
Db 2226641 GCGGCTAAAGCTCCGTCACCTGCGCTGGAATCGTGACGATGGGCTGCAGGACATGTTG 2226582
QY 2027 CGC---GTCCGATGCTGTGGGCGCAGGCGGTATCCATTGTCCTCGCTCGCGGAAGT 2083
Db 2226581 CGTAAGGCTTTTCGTCGCGCATCACCAGACCGGACACTTTGTTCGCGACACGAGGCGC 2226522
QY 2084 TTTGGTGAGGCAATTCGCGCGAGGCTTGTGGGAGCTTCTTCCGAAACCCAAAGCTGCT 2143
Db 2226521 TTACCGCGCGCAGCGCGCGCTGGATGGGCGATGTTCCCGAGAGCTGGCAGCA 2226462
QY 2144 CCCATGCTAAAGCAGGAGAGTGTATCTCTCGATAGATACCCATTGACACACCGATG 2203
Db 2226461 TCTCCGCTTGGCATGGATCGTTCTGACGGGTCTGCGACATACACCTCGACGTCCTCTC 2226402
QY 2204 TATTGGCAACGATGCGCGCTGGAATCTAGATCTCTAGCTAGACTCACAGACCGCTCGTT 2263
Db 2226401 TATTGGCAATGCTGGAACCTGGACAGTCCGATCATCGCGGAATTAACGACCGGTGAGG 2226342
QY 2264 GATGAGCAATCGAGGATTCGCGCTTAGTTAC 2297
Db 2226341 GCGGCGCAAGCGCTCTGTACCGGGCGGCAAC 2226308
```

RESULT 3  
US-09-103-840A-1/c  
; Sequence 1, Application US/09103840A  
; Patent No. 6294328  
; GENERAL INFORMATION:  
; APPLICANT: FLEISCHMAN, Robert D.  
; APPLICANT: WHITE, Owen R.  
; APPLICANT: FRASER, Claire M.  
; APPLICANT: VENTER, John C.  
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
; TUBERCULOSIS  
; FILE REFERENCE: 24366-20007.00  
; CURRENT APPLICATION NUMBER: US/09/103,840A  
; CURRENT FILING DATE: 1998-06-24  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 4411529  
; TYPE: DNA  
; ORGANISM: Mycobacterium tuberculosis  
; OTHER INFORMATION: H3/RV  
US-09-103-840A-1  
  
Query Match 5.1%; Score 121.2; DB 4; Length 4411529;  
Best Local Similarity 49.5%; Pred. No. 7.1e-24;  
Matches 433; Conservative 0; Mismatches 423; Indels 18; Gaps 4;  
  
QY 1433 CAACTGGACATTGCTCTCAATCATGATGAAGGACGCTTCGAAGCGCGCTCTTAGCC 1492  
DB 2229873 CAGCTGGCGCATGGCTGCCGTGCTGCAACTGCGAGCTTCGATGCGCGCGGAGCGC 2229814  
  
QY 1493 CTTTCATTTCCTCCCTCGGGGTGAGTCAGCGGTTAAAGCTCTCAGCATCAGTGGGT 1552  
DB 2229813 CTACATGTCACCCCGCTCGGCTGTGCTAGTCAGCGCATCAAGTCTGTGGACGACGAGTGGC 2229754  
  
QY 1553 CGAGTCTTGATCGGCGCAACCCGCAACCGGCAACGACGAGCGGTGAAGTCTCTTGTG 1612  
DB 2229753 CAGTGTCTGTTGTCAGGGAAGCATGTGCGGGCGACGCGAGGTATCCCGTGTG 2229694  
  
QY 1613 CAAGCAGCGGGAATGTTGTTGCTGCAAGCAGAACTAAAGCGCACTATCG--AC 1670  
DB 2229693 CGGTTGGCGCGCAACACGCTGCTGAGTCCGAGCGCTCGCTGAAATGGGTGGCAAC 2229634  
  
QY 1671 GCCTTCGTGA---ATCCGTTAACCATCGCATCAACGCGATTCGTTATCCACATGG 1726  
DB 2229633 CGCTCGTGAACGACGCGGATCACCATTGCGGTAAACGCGGATTCATGGCGACATGG 2229574  
  
QY 1727 TTTTCTCCCGTGTTCACAGAGGTAGCTTCTTGGGGTGGAGCAACGCTCACGCTGCGCTTG 1786  
DB 2229573 TTTTCGCGCGTGTTCAGC-----GTCTCGGCGAGTCTCTCTGACGTTCCGATC 2229523  
  
QY 1787 GAAGATGAAGCGCACATATATCTTCTGCGCGCTGGAGATGTTTTAGAGCGGTAAAC 1846  
DB 2229522 GAGGACACGAGCAATTCGCGCGGCTGCTACGCGAGGTGTGGCATGGCGCGGTGACC 2229463  
  
QY 1847 CGTGAAGCTAATCCCTGCGGAGTGAAGTAGTAGAATTTGGAACCATGCGCCACTTG 1906  
DB 2229462 ACCGAGCGGAACCGGTGCGGGCTGCGGGGTGCACCCGCTGGTGAATGCGCTACCTA 2229403  
  
QY 1907 GCATTGCAACCCCTCATTTGCGGGATGCTACATGTTGATGGGAACCTAGATTGGGCT 1966  
DB 2229402 CAGTGGCGCAGACGAGGCATTCGTGAGCGGCATCTATCGAGCGGTTCACTGCGCGCGG 2229343  
  
QY 1967 GCGATGCCCGCTTTAGCTTTCCGTTCCCAAGATGTGCTTCAAGACCGCTGACCTGGACGGG 2026  
DB 2229342 GCGGTAAAGCTCCGTCACTGCGGTGAATCTGTAGACGTGGCTGCAGGACATGTTGTG 2229283  
  
QY 2027 CGC---GTGATGTTCTGTGGGGCGCAGCGGTATTCATTGTCCTGCGCGGGAAGT 2083  
DB 2229282 CGTAGGCGCTTTCTGTCGCGCCATCCAGACGCGCATTTGTCTCCGACCAACAGAGGGC 2229223

QY 2084 TTTGCTGAGGCAATTGCGCGAGGCTTGTGGGACTTCTTCCGAAACCCAGCTGCT 2143  
DB 2229222 TTCACCGCGCAGCGCGCGGCTGGATGGGCGATGTTCCCGAGAGCTGGCAGCA 2229163  
  
QY 2144 CCATGCTAAAAGCAGGAGAGTGAATCTCTCGATGAGATACCATTTGACACACCGATG 2203  
DB 2229162 TCTCCGCTTGCCGATGATCGTTGCTAGCGGTCTGCGACATACACCTCGACGTCCTCTC 2229103  
  
QY 2204 TATTGCGCAACGATGCGCGCTGGAACTCTAGATCTCTAGCTAGACTACACAGCGCGTCTG 2263  
DB 2229102 TATTGGCAATGCTGGAACTGGACAGTCCGATATATCCGCGAATTCACGACCGTGAGG 2229043  
  
QY 2264 GATGCGCAATGCGGAGTTCGCGCTTTAGTTAC 2297  
DB 2229042 GCGCGCGCAAGCGCTGTGTACCGGGCGCAGCAAC 2229009

## RESULT 4

US-08-920-812-20/c  
; Sequence 20, Application US/08920812  
; Patent No. 5763188  
; GENERAL INFORMATION:  
; APPLICANT: Ohno, Teuneya  
; APPLICANT: Matsuhisa, Akio  
; APPLICANT: Uehara, Hirotsugu  
; APPLICANT: Eda, Soji  
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/920,812  
; FILING DATE: 29-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/362,577  
; FILING DATE: 27-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rin-Laures, Li-Hsien  
; REGISTRATION NUMBER: 33,547  
; REFERENCE/DOCKET NUMBER: 19036/32420  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5541 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli  
; STRAIN: Clinical Isolate EC-625  
; US-08-920-812-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;  
Best Local Similarity 50.5%; Pred. No. 1.7e-19;  
Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;  
  
QY 1 AGATACTCTTTGGAGAAACCATGCTAGCATTTGGTGCATTTGCTGGAAGGC 60

```

Db 3416 AATACCGCATGAGAAACCGCCTCTGCGTGGCTCATCGGTACAAAGCGGTAAAGC 3357
QY 61 TCCTTAGCTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGAGCGGCTGAGTTTCA 120
Db 3356 GCTTTATGTGCGGATCTCTCTTACTCGCCAGAGCGGACCAAAAAATGTTGAGTTGCT 3297
QY 121 GCGGAGGAGGCTGCGCGCTTCTGATTTCATCAGCAAGCTATTCCATCATTAATCGTTG 180
Db 3296 GCGGAGTGAATAATTCGCTGTTAATTCATCACTTCTGATTAATTCATCAACCGCTG 3237
QY 181 GGTGAGGAACCGCGGATGACGGTGAGAACTTTGTCAGTCACTGCGCAACATGTTCT 240
Db 3236 GGTGATATAAGCGG-----CCTGCTGATACCTCTGACCAATAAACGCGT 3192
QY 241 TGGGCTCATTTCTTCTACCACTTGGCAGGCGCTGCTCACGGACAAATATCTCATGG 300
Db 3191 GGGCTGATTGCGCTTACTCTCTGCTGAGGATTTGCTGACCGGAAATATCTCAACGG 3132
QY 301 AATTCCAGAGGTTTCCCG-----CGCCAGCGGTTAAGTCCCTGCTCGAGG 348
Db 3131 CATTCGCAAGATTCAGGATTCAGGATGATCGTGAAGGAATAAAGTTCTGAGCGGAA 3072
QY 349 CATGTTGAACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
Db 3071 AATGCTTACCGAAGCCAACTCAACAGCCTACACTTATTGAATGAATGSCACAGCAGG 3012
QY 409 CGGCGAGTCACTTGGCAGATGCGGCTTGGCATGCGGTGCTCGCGGAGCAAGGAGTACGG 468
Db 3011 TGGACATCAATGCGCAATGCGGTTAAGCTGTTGCTGAAAGA-----2967
QY 469 CGCGGATACCGTGAACAGTGCATGATTGTTGCTTCTGCTGAGTGCAGTGCAGCAACAG 528
Db 3131 CATTCGCAAGATTCAGGATTCAGGATGATCGTGAAGGAATAAAGTTCTGAGCGGAA 3072
QY 529 CTTGATTCATCAACAACTTGGAGTTTCTGAACCGCGGAGTTGGAGGCGATCGATGAGAT 588
Db 2910 CGTGAGGCGCTGAAATAATCTGACATTTAGCACCGGAGGCTGGCGGAGTATGATCAGCA 2851
QY 589 TTCCACGAGCGCGCATCAACATTTTGGGCGAAGCCCGATTCCTCAAAACCCGCGAAAA 648

```

## RESULT 5

```

US-08-920-827-20/c
; Sequence 20, Application US/08920827
; Patent No. 5770375
; GENERAL INFORMATION:
; APPLICANT: Ohno, Tsuneysa
; APPLICANT: Matsuhisa, Akio
; APPLICANT: Uehara, Hirotsugu
; APPLICANT: Eda, Soji
; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,827
; FILING DATE: 29-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/362,577
; FILING DATE: 27-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 19036/32420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5541 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: Clinical Isolate EC-625
; US-08-920-827-20

```

```

Query Match 4.2%; Score 100.8; DB 1; Length 5541;
Best Local Similarity 50.5%; Pred. No. 1.7e-19;
Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;
QY 1 AGATACTCTTTGGAGAAACCATGTACGATTGCGTGACATTTGTTGCGTCTGGAAGCG 60
Db 3416 AATATACCGCATGGAAGAAACCGCCTCTGCGTGGCTCATGCGGTACAAAGCGGTAAAGC 3357
QY 61 TCCTTAGCTGGGTATTTCTTCTACGGTCCAGAGCTCACAGCGGAGCGGCTGAGTTTCA 120
Db 3356 GCTTTATGTGCGGATCTCTCTTACTCGCCAGAGCGGACGAAAAATGTTGAGTTGCT 3297
QY 121 GCGGAGGAGGCTGCCCGCTTCTGATTTCATCAGCAAGCTATTCCATCATTAATCGTTG 180
Db 3296 GCGGAGTGAATAATTCGCTGTTAATTCATCACTTCTGATTAATTCATCAACCGCTG 3237
QY 181 GGTGAGGAACCGCGGATGACGGTGAGAACTTTGTCAGTCACTGCGCAACATGTTGCT 240
Db 3236 GGTGATATAAGCGG-----CCTGCTGATACCTCTGACCGGAAATATCTCAACGG 3192
QY 241 TGGGCTCATTTGCTTCTCACCACTTGGCAGGCGCTGCTCACGGACAAATATCTCATGG 300
Db 3191 GGGCTGATTGCGCTTACTCTCTGCTCAGGATTTGCTGACCGGAAATATCTCAACGG 3132
QY 301 AATTCCAGAGGTTTCCCG-----CGCCAGCGGTTAAGTCCCTGCTCGAGG 348
Db 3131 CATTCGCAAGATTCAGGATTCAGGATGATCGTGAAGGAATAAAGTTCTGAGCGGAA 3072
QY 349 CATGTTGAACGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408
Db 3071 AATGCTTACCGAAGCCAACTCAACAGCCTACACTTATTGAATGAATGSCACAGCAGG 3012
QY 409 CGGCGAGTCACTTGGCAGATGCGGCTTGGCATGCGGTGCTCGCGGAGCAAGGAGTACGG 468
Db 3011 TGGACATCAATGCGCAATGCGGTTAAGCTGTTGCTGAAAGA-----2967
QY 469 CGCGGATACCGTGAACAGTGCATGATTGTTGCTTCTGCTGAGTGCAGTGCAGCAACAG 528
Db 2966 ----TGATCGGTGACGTCGGTATTGATTGTTGTC--CAGCGCGCGGACAACTTGAAGAGNA 2911
QY 529 CTTGATTCATCAACAACTTGGAGTTTCTGAACCGCGGAGTTGGAGGCGATCGATGAGAT 588
Db 2910 CGTGAGGCGCTGAAATAATCTGACATTTAGCACCGGAGGCTGGCGGAGTATGATCAGCA 2851
QY 589 TTCCACGAGCGCGCATCAACATTTTGGGCGAAGCCCGATTCCTCAAAACCCGCGAAAA 648

```

Db 2850 TATCGCGGATGGGAGTGAATCTGTGGCAGGCGTCTTCGGATAAATGACCTGTGTAATA 2791  
 QY 649 CTAACCCATCAACATCAGTTTGTATGGCCATGCGGTTCATCACAACCTGCCACGACGAGTT 708  
 Db 2790 CGGGCGGAGAACCGGCCCGCATAGATCAGTGATATACGAGTCAAGTCAAGAT 2731  
 QY 709 GATCCAGCGCCACACCTTTGGGGCTGGACAGCGGCGGTGACCAATG 752  
 Db 2730 AGCCCATCAAAATGACAGAGACAAACGTCAGATGTAATG 2687

## RESULT 6

US-08-921-177-20/c  
 ; Sequence 20, Application US/08921177  
 ; Patent No. 5798211  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/921,177  
 ; FILING DATE: 29-AUG-1997  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/362,577  
 ; FILING DATE: 27-MAR-1995

ATTORNEY/AGENT INFORMATION:  
 ; NAME: Rin-Laures, Li-Hsien  
 ; REGISTRATION NUMBER: 33,547  
 ; REFERENCE/DOCKET NUMBER: 19036/32420  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312/474-6300  
 ; TELEFAX: 312/474-0448  
 ; TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5541 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: Genomic DNA  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Escherichia coli  
 ; STRAIN: Clinical Isolate EC-625

US-08-921-177-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;  
 Best Local Similarity 50.5%; Pred. No. 1.7e-19;  
 Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;

QY 1 AGATACTCTTTTGAAGAAACCATGTACGCAATTGCGTGCATTTGCGTCTGGAAGGC 60  
 Db 3416 AATACGCCGATGAGAAACCCCTCTGCGTGGCTCATGCGGTACAAGCGGTGAAGC 3357  
 QY 61 TCTTTAGTGGGTATTTCTTCCAGGTCAGAGCTCACAGCGGAGCGGCGGTGATTCAT 120  
 Db 3356 GCTTTATGCGGATCTCTCTTACTCGCCAGAGCGGACCAAAAAATGGTTGATGCT 3297

QY 121 GSCGAGGAGGGCTCCCGCTTCTGATTTCATCAGCAAGCTATTCCATCATTAATCGTGG 180  
 Db 3296 GGCAGGTGGAAAAATTCGCGTGTAAATTCATCAACCTTCGTACAAATTTACTGAACCGCTG 3237  
 QY 181 GGTGGAGGAACCGGCGGATGACGGTGTGAGAACTTTGTTCAGTCAAGTGCAGCAACAATGGTCT 240  
 Db 3236 GGTGGATAAAGCGG-----CCTGCTGGATACCTGCAAAATACGGCGT 3192  
 QY 241 TGGCGTCAATTGCTTTCTCACCATTGGCGAGGCGCTGTCCAGGACAAATATCTCGATGG 300  
 Db 3191 GGGCTGTATTGCTTTACTCTCTCGGCTCAGGGATTGCTGCCGGAATATCTCAACGG 3132  
 QY 301 AATTCACAGAGGTTCCCG-----CGCCAGCCAGGCTAAGTCCCTGTCTGAGG 348  
 Db 3131 CATTCGCAAGATTCAAGATGTCATCGTGAAGGGAATTAAGTTCTGTGCTGACCCCGAA 3072  
 QY 349 CATGTTGAACGTGAACAATATTGATATATGTCGCGAAGCTCAATGACATCGCCCGCAAGACG 408  
 Db 3071 AATGCTTACCGAAGCAACCTCAACAGCCTACACTTATTGAATGAATGSCACAGCAGCG 3012  
 QY 409 CGGGCAGTCACTTGGCGAGATGGCGCTTGCATGGGTGCTGCGCGAGCAAGGAGATACGG 468  
 Db 3011 TGGACAATCAATGGCGCAATGGCGCTTAAGCTGTGCTGCTGAAAGA-----2967  
 QY 469 CGCGGATACCGTGACCACTGCAATGATTGCTGCTGCTGCTGAGCAGTGGACAGTGGACACAG 528  
 Db 2966 ----TGATCGGTGACGTCGGTATTGATTTGTTGCTG-CAGCCGCGGCAACTGTGAGGAA 2911  
 QY 529 CTTGATTCACTCAACAACCTTGGAGTTTCTGACCCCGAGTTGGAGCGCATCGATGAGAT 588  
 Db 2910 CTTGAGGCGCTGATATCTGACATTTAGCACCGAGAGCTGGCGCAGATGATCACA 2851  
 QY 589 TTCCACGACGCGCGCATCAACATTTGGGCGAAGCCACCGATTCCTCAAAACCCCGCAAAA 648  
 Db 2850 TATCGCGATGGCGAGCTGAATCTGTGGCAGCGCTTCCGATAAATGACCTGTTAATAA 2791  
 QY 649 CTAACCCATCAACATCAGTTTGTGATGGCCANTGCGGTCTATCAACTGCCACGACAGCTT 708  
 Db 2790 CGGGCGGAGAACCGGCCCGCATTAAGATCAGTGAATGATGATGAGTCAAGTCAAGT 2731  
 QY 709 GATCCAGCGCCACACCTTTGGGGCTGGACAGCGGCGGTGACCAATG 752  
 Db 2730 AGCCCATCAAAATGACAGAGACAAACGTCAGATGTAATG 2687

## RESULT 7

US-08-362-577C-20/c  
 ; Sequence 20, Application US/08362577C  
 ; Patent No. 5807673  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ohno, Tsuneya  
 ; APPLICANT: Matsuhisa, Akio  
 ; APPLICANT: Uehara, Hirotsugu  
 ; APPLICANT: Eda, Soji  
 ; TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 ; NUMBER OF SEQUENCES: 25  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 ; STREET: 6300 Sears Tower, 233 South Wacker Drive  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: United States of America  
 ; ZIP: 60606-6402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/362,577C  
 ; FILING DATE: 27-MAR-1995  
 ; CLASSIFICATION: 536



ATTORNEY/AGENT INFORMATION:  
 NAME: Rin-Laures, Li-Hsien  
 REGISTRATION NUMBER: 33,547  
 REFERENCE/DOCKET NUMBER: 19036/32420  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5541 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: Clinical Isolate EC-625  
 US-08-362-577C-20

Query Match 4.2%; Score 100.8; DB 1; Length 5541;  
 Best Local Similarity 50.5%; Pred. No. 1.7e-19;  
 Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;

QY 1 AGATACCTCTTTCGAGAAACCATGATGCGTTCGATGACATTTGTTGGCTCTGGAAGGC 60  
 DB 3416 AATACGCCGATGAGAAACCGCTCTGCGTGGCTATGCGGTACAAAGCGGTAGGC 3357

QY 61 TCTTTACGTGGGTATTCTTCTACGGTCCAGAGCTCACAGCGGAGCGGCTGAGTTTCA 120  
 DB 3356 GCTTTATGTGCGGATCTCTCTTACTCGCCAGAGCGGACGCAAAAATGTTGAGTTGCT 3297

QY 121 GCGGAGGAGGCGTCCCGCTTCTGATTCATCATCAGCAAGCTATTCATCATTAATCGTTG 180  
 DB 3296 GCGGAGTGGAAAATTCGCTGTTAATTCATCACTTCGTGACAAATTTACTGAACCGCTG 3237

QY 181 GGTGAGGAACCGGCGGATGACGGTGAGAACTTTGTGAGTCAGCTGCGCAACAATGGTCT 240  
 DB 3236 GGTGATAAAGCGG-----CCTGCTGGATACCTTGCAAAATAACGGCT 3192

QY 181 GGTGAGGAACCGGCGGATGACGGTGAGAACTTTGTGAGTCAGCTGCGCAACAATGGTCT 240  
 DB 3236 GGTGATAAAGCGG-----CCTGCTGGATACCTTGCAAAATAACGGCT 3192

QY 241 TGGCGTATGTTTCTACCACTTGGCGAGCGCTGCTCACGGAACAATATCTCGATGG 300  
 DB 3191 GGGCTGTATTGCTTTACTCTCTGCTCAGGATGCTGACCGGAAATATCTCAACGG 3132

QY 301 AATTCAGAGGTTCCG-----CGCCAGCGGGTAAGTCCGTCTGAGGG 348  
 DB 3131 CATTCGCAAGATTACGGATGATCGTGAAGGGAATAAAGTTCTGCTGACGCCGAA 3072

QY 349 CATGTTGAACGTGAACAATATTGATGTCGCAAGCTCAATGACATCGCCAGGAACG 408  
 DB 3071 AATGCTTACGAGCACTCAACAGCTTACACTTATTGATGAATGACACAGCAGCG 3012

QY 409 CGGCGAGTCACTTTCGCGAGATGCGCTTGCATGGGTGCTCGCGGAGCAAGAGGATACGG 468  
 DB 3011 TGGACAATCAATCGCGAAATGCGTTAAGCTGTTGCTGAAAGA-----2967

QY 469 CGCGGATACGTCACCAAGTCATGATGTTGCTGCTGATGAGCTGACAGCAACAG 528  
 DB 2966 ---TGATCGGTCAGCTGCGTATGATGTTGTC- CAGCCGCGCGGCAACTTGAAGAGAA 2911

QY 529 CTTGATTCTCAACAACCTTGAGTCTTCTGACGCGGAGTTGGAGCGGATCGATCAGAT 588  
 DB 2910 CGTGCAGGCGCTGATTAATCTGATTTAGCACCGAGGAGCTGGCCGATGATCAGCA 2851

QY 589 TTCCGAGGCGCGCATCAACATTTGGGGAAGGCGACCGATTCCAAAACCGCGAATA 648  
 DB 2850 TATCGCGGATGGGAGCTGAATCTGTGGCAGGCGTCTTCGATAAATGACCTGTTAATA 2791

QY 649 CTAACCCATCAACATCAGTTTGTATGCCAATGCGGTTCATCACTGCCACGACGCTT 708  
 DB 2790 CGGGCGGAGAACCGGCCCGCATATAGATCAGTGATTTATGACAGATCGATCGGTCAAGAT 2731

QY 709 GATCCAGCGCCACACCTTGGGCTGGACAGCGGCGGTGACAATG 752

DB 2730 AGCCCATCAAAATGCGAGGAGCACAAACGTGATGATATG 2687

RESULT 8  
 US-08-920-828-20/c  
 Sequence 20, Application US/08920828  
 Patent No 585398  
 GENERAL INFORMATION:  
 APPLICANT: Ohno, Tsuneya  
 APPLICANT: Matsuhisa, Akio  
 APPLICANT: Uehara, Hirotsugu  
 APPLICANT: Eda, Soji  
 TITLE OF INVENTION: Probe for Diagnosing Infectious Disease  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
 STREET: 6300 Sears Tower, 233 South Wacker Drive  
 CITY: Chicago  
 STATE: Illinois  
 COUNTRY: United States of America  
 ZIP: 60606-6402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/920,828  
 FILING DATE: 29-AUG-1997  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/362,577  
 FILING DATE: 27-MAR-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Rin-Laures, Li-Hsien  
 REGISTRATION NUMBER: 33,547  
 REFERENCE/DOCKET NUMBER: 19036/32420  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 312/474-6300  
 TELEFAX: 312/474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 5541 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: Genomic DNA  
 ORIGINAL SOURCE:  
 ORGANISM: Escherichia coli  
 STRAIN: Clinical Isolate EC-625  
 US-08-920-828-20

Query Match 4.2%; Score 100.8; DB 2; Length 5541;  
 Best Local Similarity 50.5%; Pred. No. 1.7e-13;  
 Matches 386; Conservative 0; Mismatches 332; Indels 46; Gaps 4;

QY 1 AGATACCTCTTTCGAGAAACCATGATGCGTTCGATGACATTTGTTGGCTCTGGAAGGC 60  
 DB 3416 AATACGCCGATGAGAAACCGCTCTGCGTGGCTATGCGGTACAAAGCGGTAGGC 3357

QY 61 TCTTTACGTGGGTATTCTTCTACGGTCCAGAGCTCACAGCGGAGCGGCTGAGTTTCA 120  
 DB 3356 GCTTTATGTGCGGATCTCTCTTACTCGCCAGAGCGGACGCAAAAATGTTGAGTTGCT 3297

QY 121 GCGGAGGAGGCGTCCCGCTTCTGATTCATCATCAGCAAGCTATTCATCATTAATCGTTG 180  
 DB 3296 GCGGAGTGGAAAATTCGCTGTTAATTCATCACTTCGTGACAAATTTACTGAACCGCTG 3237

QY 181 GGTGAGGAACCGGCGGATGACGGTGAGAACTTTGTGAGTCAGCTGCGCAACAATGGTCT 240  
 DB 3236 GGTGATAAAGCGG-----CCTGCTGGATACCTTGCAAAATAACGGCT 3192



Query Match 1.7%; Score 39.6; DB 2; Length 1935;  
 Best Local Similarity 49.1%; Pred. No. 0.14;  
 Matches 105; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 520 GGACAAACAGCTTGATTCACTCAACAACTTGGAGTTTCTGACGCGAGTTGGAGCGGAT 579  
 |||||  
 DB 1263 GTCACCGGATCGGTTCGCCATCAGTACGCTGTGTGGCCCTCAAGTGGAGTCTT 1322  
 |||||

QY 580 CGATGAGATTTCCACGAGCGCCGCGATCAACATTTGGGGGAAGGCCACCGATTCCAAAAC 639  
 |||||  
 DB 1323 CGTGGTCATTGGCCATAGCTGCTGGGTGCGATCAGGGGCTCTCTCCCTCCAGGAAC 1382  
 |||||

QY 640 CCGCGAAACTAACCATCAATCAGTTGATGGCCATGCGGTTCATCAACTGCCAC 699  
 |||||  
 DB 1383 CCGCGAGCTTACACCTTCCTTCGAGGACTGGGTAAAGATCGGCTTCATTGCCAA 1442  
 |||||

QY 700 GACGAGCTTGATCAGCGCCGACACCTTGGGGCTG 733  
 |||||  
 DB 1443 GATGAGGTTAAGAAAGACGACGCTCGGTGCGG 1476  
 |||||

RESULT 11  
 US-09-103-840A-2  
 ; Sequence 2, Application US/09103840A  
 ; Patent No. 6294328  
 ; GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 4403765  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; FEATURE:  
 ; OTHER INFORMATION: CDC 1551  
 ; OTHER INFORMATION: "n" bases at various positions throughout the sequence  
 ; OTHER INFORMATION: represent a, t, c or g  
 US-09-103-840A-2

Query Match 1.6%; Score 37.6; DB 4; Length 4403765;  
 Best Local Similarity 47.5%; Pred. No. 34;  
 Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 695 GCACAGCAGCTTGATCCAGCGCCACACCTTGGGGCTGACAGCGGGGTGACAATGCT 754  
 |||||  
 DB 720208 GCCATCAGCGCGTAAACCCGTGCAACGGTTGTGGCCGCGACCGCGGATCTCCCG 720267  
 |||||

QY 755 GCTGCGCGAAACCCAGCGGAGACCATCAGCTTGCCTGCGGAAACCGCGCGCGG 814  
 |||||  
 DB 720268 GCGCGGAAGAAACCGGCGAGGAATCCCGCCAGCAGGTCTCTCGATCTCGACGGCTG 720327  
 |||||

QY 815 AAATCCACCGTCCGGTGTCCGGTATTGGCGCGGACCGCGCGGATAACACAAACGG 874  
 |||||  
 DB 720328 TGGTCGTGACCCCGAACATTCGTGTCGCGGCCGTTGCAAGTGACAGGCCACCG 720387  
 |||||

QY 875 TCCAAATACGATTCGGGTTCAACACAGGTGACGACGATTGCCATCAACATGGCTT 930  
 |||||  
 DB 720388 ACCGGGGGCCCGGCGAGCTCCGCGCGCGCTCGACGCGCAGGCGGAGTCTT 720443  
 |||||

RESULT 12  
 US-09-103-840A-1  
 ; Sequence 1, Application US/09103840A  
 ; Patent No. 6294328

GENERAL INFORMATION:  
 ; APPLICANT: FLEISCHMAN, Robert D.  
 ; APPLICANT: WHITE, Owen R.  
 ; APPLICANT: FRASER, Claire M.  
 ; APPLICANT: VENTER, John C.  
 ; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
 ; TITLE OF INVENTION: TUBERCULOSIS  
 ; FILE REFERENCE: 24366-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/103,840A  
 ; CURRENT FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 2  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 4411529  
 ; TYPE: DNA  
 ; ORGANISM: Mycobacterium tuberculosis  
 ; OTHER INFORMATION: H37RV  
 US-09-103-840A-1

Query Match 1.6%; Score 37.6; DB 4; Length 4411529;  
 Best Local Similarity 47.5%; Pred. No. 34;  
 Matches 112; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 695 GCACAGCAGCTTGATCCAGCGCCACACCTTGGGGCTGACAGCGGGGTGACAATGCT 754  
 |||||  
 DB 718775 GCCATCAGCGCGTAAACCCGTGCAACGGTTGTGGCCGCGACCGCGCGATCTCCCG 718834  
 |||||

QY 755 GCTGCGCGAAACCCAGCGGAGACCATCAGCTTGGCGGAAACCGCGGACCGCGCGG 814  
 |||||  
 DB 718835 GCGCGGAAGAAACCGGCGAGGAATCCCGCCAGCAGGTCTCTGATCTCGACGCTCG 718894  
 |||||

QY 815 AAATCCACCGTCCGGTGTCCGGTATTGGCGCGGACCGCGCGGATAACACAAACGG 874  
 |||||  
 DB 718895 TGGTCGTGACCCCGAACATTCGTGTCGCGGCCGTTGCAAGTGACAGGCCACCG 718954  
 |||||

QY 875 TCCAAATACGATTCGGGTTCAACACAGGTGACGACGATTGCCATCAACATGGCTT 930  
 |||||  
 DB 718955 ACCGGGGGCCCGGCGAGCTCCGCGCGCGCTCGACGCGCAGGCGGAGTCTT 719010  
 |||||

RESULT 13  
 US-08-358-117-1/c  
 ; Sequence 1, Application US/08358117  
 ; Patent No. 5608147  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Kaphammer, Bryan J.  
 ; TITLE OF INVENTION: tfda Gene Selectable Markers in Plants and the  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESS: Steine, Kessler, Goldstein & Fox  
 ; STREET: 1100 New York Avenue, Suite 600  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20005  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/358,117  
 ; FILING DATE: 15-DEC-1994  
 ; CLASSIFICATION: 800  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Esmond, Robert W.  
 ; REGISTRATION NUMBER: 32,893  
 ; REFERENCE/DOCKET NUMBER: 1405.0030001  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (202) 371-2600  
 ; TELEFAX: (202) 371-2540  
 ; INFORMATION FOR SEQ ID NO: 1:



